AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 18, NUMBER 2 FEBRUARY, 1950

Physics in Industry*

C. G. SUITS
General Electric Company, Schenectady, New York

TWENTY-FIVE years ago, the subject upon which I am to address you would have afforded but meager fare—and fifty years ago, it would have been completely meaningless. Indeed, the entire history of this subject is encompassed by the careers of men now living. It is true that since the industrial revolution, the roots of industry have had their beginnings in a fertile substratum of scientific exploration, but it is only since the turn of the century that the vitalizing power of scientific research has been consciously and effectively directed toward industrial progress. In this half-century, industry has undergone a significant mutation from a primitive system based upon ancient arts, manual skills, and simple inventiveness to a complex machine constructed by the advances of science and lubricated by the refined processes of modern technology.

I was reminded by Karl Darrow, when he was developing the extremely dubious thesis that I was in a better position than he, to talk on this subject, that the birth of the American Physical Society in 1899 nearly coincided with the founding of the General Electric Research Laboratory in Schenectady in 1900 by Dr. W. R. Whitney. These events were undoubtedly unrelated, except that their inception was probably symptomatic of the renascence of science which followed the far reaching discoveries of this period around the beginning of the 20th century. In 1896 Roentgen had observed the penetrating rays which bear his name, and Becquerel had found a peculiar emanation which originated in uranium. At about the same time Marconi was performing some startling experiments with electric waves following Hertz' work a few years previously. Not long afterwards the Curies discovered radium and its progeny, and some of its effects could be observed in a cloud chamber which had just been invented by C. T. R. Wilson. In 1897 J. J. Thomson solved the mystery of the nature of the cathode rays, and electronics was off to a modest beginning. In 1900 Max Planck published his first paper on the quantum theory. This was the exciting historical environment in which the American Physical Society was born; industrial scientific research, originally principally physical research, made an inauspicious start at about this same time.

Since this may be done without embarrassment, and indeed with a good deal of satisfaction, it is of interest to consider briefly the place which physics and physicists have attained in industry today. It is not necessary to remind you of the manifold applications of the principles of physics to industrial progress, for the contributions of physics are well known to everyone. The applications of physics to the development of new products, new machines, and new processes, to the complex field of instrument design, and to quality control and standardization are too numerous to mention. However, as I considered the probable nature of your interest in

* Address to the American Physical Society, Cambridge, Massachusetts, June 17, 1949.

this subject, a number of pertinent questions came to mind.

Is research in physics really supported to any appreciable extent by American industry? What physical facilities for research work are available in industrial organizations? What kind of research problems are being actively pursued in industrial laboratories? What is the status of the physicist in industry—does he have an opportunity to develop his own particular interests in research, and is he free to publish and to engage in scientific intercourse with scientists in other institutions? What is his opportunity for advancement, and towards what does he advance? What is the attitude of industry toward research of a fundamental nature?

In an endeavor to formulate partial answers to these questions, I have consulted with the directors of research in about 30 leading American industries, and this is a good time to acknowledge their hearty cooperation and assistance in collecting material for this talk. While it is difficult to obtain reliable national figures on the capital investments in research facilities, the annual expenditures for research, and the number of scientists of various kinds employed in industrial laboratories, I can assure you that industry is supporting physical research on an unprecedented and ever increasing scale. The replies to my inquiries indicate a very broad range of interests and a keen awareness of the prime importance of research to industrial progress.

To illustrate the attitude of one segment of American industry on the subject of scientific research, I can do no better than quote from a talk by Mr. Charles E. Wilson, president of my company, made recently on this subject. He said, "I believe that industry has a duty, to itself and to the society in which it lives, to undertake, encourage, and maintain extensively both fundamental and applied research." I can tell you that he and his organization have given convincing and concrete evidence of their sincere and earnest belief in this policy.

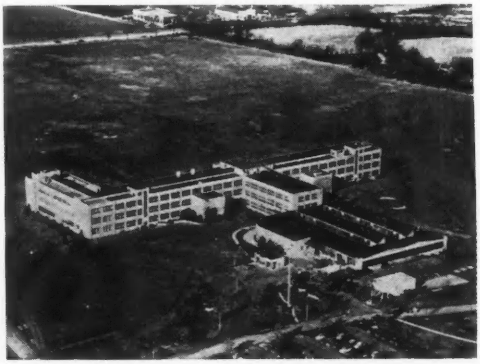

FIG. 1. RCA Laboratories, Princeton, New Jersey.

The expenditures for industrial research have quadrupled in the last two decades. In 1930, the industrial support of research amounted to 116 million dollars out of a total for the country of 166 million dollars. By 1940 the figure had doubled—234 million out of a total of 345 million. It is estimated that by 1950, it will have more than doubled again, the estimated figure being 500 million out of a total of 700 million. These figures, taken from the report of the President's Committee on Research and Development (Steelman report), show that for the past twenty years, industry has been contributing an average of 70 percent of the national investment in research.

Dr. M. H. Trytten, Director of the Office of Scientific Personnel, tells me that about one-third of the 9000 physicists classified as professional in this country are employed in industry. Dr. Henry Barton, Director of the American Institute of Physics, estimates that within 10 years industry will be claiming 70 percent—over two-thirds—of the physicists.

As regards physical facilities and equipment, I think it is safe to say that industrial laboratories now offer facilities unexcelled anywhere in the world. If there were some lingering doubts, many years ago, concerning industry's interest in research, these have been dissipated completely by the enthusiasm with which the major industrial organizations have invested hundreds of millions of dollars in new laboratories and research equipment. This is naturally not a result of any sentimental infatuation with science, but rather it reflects a firm conviction that the clink of dollars spent for research eventually makes a noise which will be pleasant to the ears of stockholders, to mention just one important but largely neglected segment of our economy.

Although some of you are well acquainted with some of the industrial research laboratories, many of these laboratories, particularly the ones that have been completed recently, may not be familiar to you. To add perspective to what I wish to say, I felt it would be helpful to show you photographs of three typical research laboratories in industry.

The first unit of the RCA Laboratories, shown in Fig. 1, was completed at Princeton, New Jersey in 1942; a new wing was added recently. The laboratory includes a model shop and a fine technical library. In determining the location of this new laboratory, RCA decided against placing it in the middle of one of its large manufacturing plants, in favor of a site which was near the center of gravity of its entire operation comprising many plants and related technical activities.

One of the largest laboratories in the construction materials industry is the recently completed laboratory of Johns-Manville Company at Manville, New Jersey, shown in Fig. 2. One side of the central corridor of the main building includes facilities for general laboratory work, while the opposite side is given over completely to a high bay experimental space suitable for pilot equipment development which is carried on in parallel with laboratory studies.

Figure 3 shows the new General Electric Research Laboratory at the Knolls, near Schenectady, New York. The first unit was occupied last fall; an additional unit is in construction.

One cannot help but be impressed by the fundamental and increasingly important role which research is playing in American industry, and the outstanding plant facilities which are being provided in support of this role.

Buildings and facilities and equipment—the tools of research—are of course important, but more important than these are the people, and what they do, the kinds of policies that determine the environment in which they work, and the satisfaction to which they may aspire as scientists. It is this side of industrial research in physics to which I want to speak next.

What kind of research problems are being investigated in industrial laboratories? A casual survey of the fields of activity of this representative group of industrial laboratories reveals that practically every branch of physics is well represented. It would be futile for me to attempt here to give anything like a comprehensive survey of this subject, but a few significant examples will illustrate the diversity of industrial research.

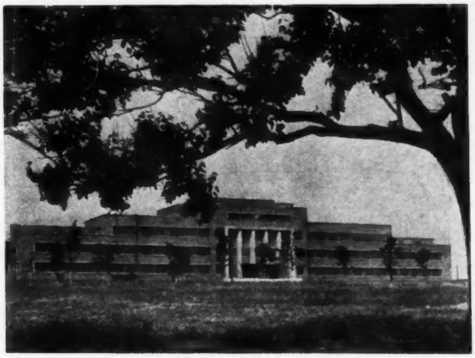

FIG. 2. Johns Manville Company, Research Laboratory, Manville, New Jersey.

The petroleum industry is a good example of an industry that actively pursues investigations along a number of special lines in physics. Analytical methods involving particle bombardment, mass spectrometry, infra-red spectroscopy, electron microscopy, electron and x-ray diffraction are in common use.

Magnetic surveying is being studied, based upon observations relating polarity, susceptibility, and other measurable properties of core samples or exposed strata to geological history and the composition of the earth's crust. Oil physicists have been responsible for extraordinary progress in the measurement of gravity. Some recent work in this field has been reported upon by Woollard of the Sun Oil Company, who made an 82-thousand mile trip correlating gravity reference bases over a much larger area than had been previously covered. From these impressive mileage statistics, it may be inferred that here at last is a type of research which may compete in geographical glamour even with cosmic-ray studies.

Along the line of mathematical research, extensive studies are aimed at projecting magnetic fields and gravity fields measured at the earth's surface to elevations above and below the level of observation.

Explorations of the penetration of the earth

FIG. 3. General Electric Research Laboratory at The Knolls, Schenectady, New York.

by ultra high frequency radiation are being undertaken as a possible new method of prospecting for oil and metallic minerals. Also, extremely low frequency electrical currents are employed in a system known as "Elflex." Another set of electrical studies of the phenomena known as *telluric currents* deals with the currents that flow on a vast scale through the crust of the earth, superinduced by extraterrestrial influences.

Seismic investigations are concerned with ground-wave patterns, electrodynamic and magnetostrictive detectors, and the transfer of energy between air waves and water waves and the earth. The determination of the depth of fluid levels involves acoustical, electrical, and mechanical methods, based upon theoretical studies and experimental procedures. Extensive research in all of the aforementioned fields has been done by the Standard Oil Company of New Jersey, the Gulf Oil Company, and the Sun Oil Company. Figure 4 shows the Gulf Oil Company airborne magnetometer trailing the airplane at the end of a cable. The measurement of gravity has a long industrial history with the oil companies and has led to some outstanding instrument development.

As a second example of an industry in which physicists are making important contributions, let us take a look at the rubber industry. Physical research in the kinetic theory of elasticity is important to an understanding of the mechanical behavior of high polymers and to the full utilization of the chemical process of vulcanization.

Special methods are required for the determination of the molecular weights of the large chain molecules of natural or synthetic elastomers. Optical scattering from dilute solutions has been developed both theoretically and experimentally so that it is possible to determine not only the molecular weight, but also the spacial extension of the molecules while in solution. Infra-red spectroscopy and x-ray diffraction methods are being increasingly developed and applied to problems in the structure and behavior of elastomers. Mechanical hysteresis studies are important, and turn out to involve considerably more than elementary mechanics, as a perusal of current articles on the dynamic properties of rubber will reveal.

The electrical industry, including the vast and ever expanding field of communication, is actively concerned with a wide variety of research problems in physics. The physics of the solid state is an especially vital field that bears upon a great number of specific problems, such as piezoelectricity, superconductivity, ferro-electrical effects, electro-optical properties, and semiconductors. This latter class of materials, best represented by silicon and germanium, is of importance in new types of detectors, amplifiers, and oscillators, of which the transistor developed at the Bell Telephone Laboratories is a noteworthy example. The relatively new field of physical metallurgy is concerned with the application of solid state physics to fundamental studies of metallurgical properties—creep, fracture, fatigue, age hardening, heat treatment, and the development of new alloys for special purposes.

The phenomenal rise of television as a brand new multimillion dollar industry almost overnight is a great satisfaction, as well as a considerable relief, to the industrial laboratories that have for years been pouring millions of dollars down what was occasionally referred to as "that rat hole." By *clairvoyance after the fact* it now becomes abundantly clear that as a field of endeavor for the physicist, television is as fertile with potentialities as was radio twenty-five years ago. As the present advanced state of the radio art is the consequence of intensive research—both fundamental and applied—so the development of this new art depends upon the contributions of research and affords important opportunities for the physicist.

In spite of its long history, that segment of physics generally included in the term elec-

tronics continues to present important opportunities for investigation, especially in gas discharges, secondary emission, space charge wave phenomena, electron optics and ultra high frequency generators and their specialized circuitry. The increasing importance of ultra high frequencies has raised innumerable problems regarding the generation and transmission of microwaves, the dielectric properties of matter, and resonant systems in general.

A field of obvious interest to the physicist and possibly great future importance to the country is the development of atomic power. Under the support of the AEC, several companies are actively engaged in research on this scientific frontier. Before nuclear energy becomes a source of useful power, a vast amount of physical research and technological development must be done, unfortunately at very great cost. This very great cost will not be due to "incredible mismanagement,"[1] but to a fact of nature which I would like to point out.

The electronics industry grew from small beginnings, because the small beginnings had a market value and the profits from their sale supported further research which led to more marketable products and so forth. The development of atomic power differs from electronics fundamentally in its laboratory-to-market pattern. To develop atomic power it is necessary to build experimental power piles as pilot plants from which to learn enough to build the full scale plants of the future. Unfortunately, a tiny pile will not react, so we have to build a big one as the very first experiment, and even when it has been stripped of all of its chrome plating and been placed in the most temporary kind of a structure, it costs a great many millions of dollars. And, at best, it is only a first experiment in a series of experimental power plants, each of which becomes progressively more expensive. Because of this circumstance of nature, there are no private resources in the country which can finance the development of atomic power. Risk capital is required on a scale that can be countenanced only by the national government. The Argonne National Laboratory, the Westinghouse Company, and General Electric all have substantial projects underway in this field under AEC sponsorship. The first phase of this development, leading to the building of the first experimental power plants, involves many difficult problems, principally in physics and mechanical engineering. In the second and third phases, many years hence, involving larger and more practical plants, the chemists will occupy the center of the stage. I hope they learn their lines and perform well, because they have a vitally important and technically fascinating part to play. This is a scientific frontier of great challenge. Fortunately, the importance of the work is matched, if not exceeded by its inherent scientific interest, so that some very able people are engaged in the work. This is the best assurance that can be given that it will have a successful outcome.

Many of the larger industrial laboratories carry out investigations that have no immediate bearing on developments related to their business interests, but may be purely scientific exploration in new fields. Some of these projects are, of course, undertaken at the request of the government under Federal Contracts, but many of them have their genesis in the interests and curiosities of individual members of the research staffs.

In our laboratory at Schenectady we are at the present time engaged in an extensive program in meteorological research which is properly in this category. This work, which was initiated by Langmuir and Schaefer as a part of our exploratory research program is now supported by

[1] An accusation made during a Congressional investigation of the AEC early in 1949.

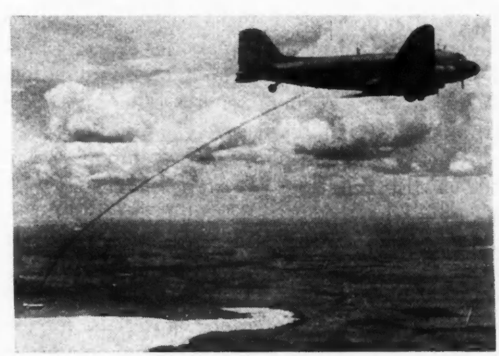

FIG. 4. Gulf Oil Company: airborne magnetometer trailing from the airplane at the end of a cable.

government contract. We are finding that the theory of nucleation which underlies this phenomenon in its various aspects is applicable to a variety of problems in many fields.

Another activity of our laboratory (entirely apart from the work of the Knolls Atomic Power Laboratory) is research in nuclear physics. A group of physicists has been engaged for three years in the design and construction of a 300-Mev non-ferromagnetic synchrotron which will be used for nuclear research. This is the culmination of a long series of high energy devices which progressed from million-volt x-ray equipment through betatrons, iron core synchrotrons, biased betatrons, and now to a high energy machine with an air core.

The aircraft industry is a fertile field for research in aerodynamics and thermodynamics. The development of supersonic planes, guided missiles, the all-wing structure of which the Northrup Flying Wing is the spectacular pioneer, and power plants of new design—these all demand the type of research for which the physicist is best qualified. The problem of propulsion, involving such radical developments as the turbo-propeller, the turbo-jet, and the rocket affords an exciting field of endeavor.

The manufacture of glass, one of our oldest industries, founded on primitive arts, is now one that draws extensively upon physical research. The ever-increasing variety of special glasses designed for specific purposes and made to rigid specifications calls for a wide range of studies correlating theory and practice. The formulation of basic laws relating the optical, electrical, and physical properties of glass to its composition and manufacturing treatment involves the application of physical methods of investigation. A recent development of the Corning Glass Company is a photosensitive glass.[2]

A field of investigation of great importance to the electrical industry is growth of crystals with piezoelectric properties. Leading work in this field has been done by the Brush Development Company and by the Bell Telephone Laboratories. Work is also being done on the piezoelectric properties of ceramic materials, for example, barium titanate. These ceramic materials exhibit interesting ferro-electric properties that are closely akin to the ferromagnetic properties of permanent magnetic materials.

It would be remiss to neglect to refer to the significant field of photography. Not only is this an art which makes intimate contact with all branches of science, but it is a fertile field of research in its own right. The science of sensitometry, the study of the physical nature of the latent image, the reproduction of color, and the application of photographic methods to the detection of nuclear particles are but a few of the intriguing fields of investigation in photography. Eastman Kodak Company has played a leading role in research in this field and in adapting the techniques of photography to the problems of science and industry. At least one type of particle, the heavy π-meson, was discovered by means of the photographic plate.

The few illustrations I have drawn are but representative examples which could be duplicated many times over. The steel, aluminum, paper, glass, chemical, and aircraft industries also afford fertile fields for research—both fundamental and applied.

And now what about the status of the industrial physicist? I believe it can be said truthfully that the opportunity in industry for the physicist to find problems that make the best use of his training and interests compare favorably with those to be found in academic laboratories. In the larger industrial laboratories, especially, the diversified activity affords a wide range of choice for the free play of individual talent.

The right to publish the results of his research in his own name and to confer freely with other workers in his own and related fields is an important consideration to the research scientist and is now common practice in industrial laboratories. It is often said that the patent situation is a bar to the exchange of scientific information. Some thought will show that this is a mistaken notion. The patent structure when properly employed by alert and active legal departments, rather than retarding the release of information, serves to expedite the prompt publication of results.

With regard to his membership in scientific

[2] Riess, Bosch, and Reboul, *Am. J. Physics* **16**, 399 (1948).

societies, his attendance at and participation in the meetings of these societies, his general intercourse with other scientists, and his opportunities to publish technical books and articles, the industrial scientist is on a par with any other.

And finally, what is the attitude of industry toward research of a fundamental nature? It must, of course, be admitted that only the larger laboratories can afford to support fundamental research to any great extent. In some few laboratories, as much as fifty percent of the effort is devoted to the search for new knowledge without any immediate application in mind. A large part of the research of industrial laboratories is devoted to long-range investigations on problems related to the interests of the company—so-called applied research. One fact that stood out among the replies that I received from various industrial laboratories was a keen awareness of the essential nature of research and the necessity for the long-range viewpoint. No longer does the director of industrial research expect the answer to a scientific problem the day after tomorrow. This is a wholesome situation, and one that makes for an environment conducive to the true spirit of research.

I am glad to report that in our experience the problems of security and classified work have not turned out to be as difficult as everyone anticipated immediately following the war. I have not heard anyone denounce security regulations for some time, at least not in a ringing voice; I would not pretend that anyone likes these regulations, but the present necessity for them in our troubled world is in no doubt. If the classified work is of real fundamental significance and challenge; for example, the development of atomic power, top grade scientists will overlook the disadvantages of security. Although the restrictions on the publication of classified work are not entirely satisfactory, we can live with them until we learn how to handle this problem better.

I have one final plea to scientists going into industrial research. In industry we have gone a long way in making our buildings and facilities versatile. We can move partitions and laboratory services with great freedom and dispatch, and we have great ability in building instruments and equipment and pilot plants for almost any type of investigation; but we have great need for corresponding flexibility in the training of our scientists. There will, of course, always be a need for physicists who are highly specialized in a narrow field of investigation, but in addition to these we have a great need for people who by disposition and broad training are highly versatile in their scientific abilities. These individuals can, with suitable apprenticeship, become progressively expert in many fields, as these fields become important in the ever-changing picture of scientific research. In short, we need more *general practitioners of physics*. To fill this specification takes one part of training—fundamental, broad training—and one part of courage. For it does take a lot of courage to leave a field in which one has become an expert, for a new field in which one will be a neophyte. But the new field is a new challenge and will provide great stimulus to further accomplishment.

Even the hardest heads are aware that research is profitable and necessary, and fail to encourage it because they lack the means rather than the will. . . . On one point I am quite clear, the public will expect that any industry that is protected by a tariff should be thoroughly aware of the need of organization and of scientific research.—E. RUTHERFORD (1932).

New Application: Electronic Time-Delay Circuit

I. Clyde Cornog
University of Pennsylvania, Philadelphia 4, Pennsylvania

THREE years ago I described in this Journal[1] an electronic circuit known as a "time-delay" circuit, giving several common but not very well-known applications of the circuit, such as regulating the time sequence of operations in a Wilson cloud chamber, and one or two better known ones, such as timing of photographic exposure during enlarging, etc.[2] At that time it was stated that I had used this circuit as a basic part of a device for cutting off the advertising razzle-dazzle to which otherwise one must listen whenever the radio is in operation, merely by a handclap or other sharp sound. This device, "Advertiser Killer," or "*A-K*," as I call it privately, has functioned so satisfactorily during several years of constant use, and has excited so much amused comment, that I thought others might like to read a description of it.

Figure 1 presents the details of the *A-K* circuit in the normal state, radio sounding. The circuit may be considered as consisting of four parts; *viz.*, the microphone with its amplifying system at the left, the gas triode 885 "trigger" tube in the middle, the time-delay circuit shown in heavy lines at the right, and the power supply below. A horn of some sort collects the sound energy from the handclap and feeds it into the microphone *M*. For constancy a flashlight dry cell is used in this branch of the circuit, and the circuit is broken by relay *V* when the device is turned off. The output from the microphone circuit, amplified as shown, is fed to the grid of the 885 tube, causing relay *X* to close. When *X* closes, the time-delay circuit operates, causing relay *Z* to fail. When *Z* fails two events occur: the plate circuit of 885 is broken, thus giving its grid control again, and the line to the radio speaker is broken, bring about those 40 or 50 seconds of quiet in the nick of time. The graph on the circuit diagram shows how the time-delay interval varies with the numbers on the dial of variable resistor *R*. The dial divisions are equal, and the resistor is a carbon potentiometer, Mallory UC504, 3 megohms, No. 1 taper.

In the circuit shown, the three relays are ordinary plate circuit relays, but generally speaking their operation characteristics must be chosen to suit the power supply. With the power supply shown, relays *X* and *Z* need not be very sensitive, operating, say, at 15 ma; but relay *V* must operate at about 5 ma since it must stay closed at all times whatever the current fluctuation, and since the current in this branch of the circuit falls to about 5 ma when thyratron 2A4G conducts. There are only a few items in the circuit which are critical in value. In the time-delay branch the delay interval depends, of course, upon *C*, *R* and *S*, and upon the charging voltage. Also, the resistor (100,000 ohms) in the charging circuit of *C* must be large in order to lower the drain on the power supply. It will be noted that current flows through this resistor continuously. In the circuit diagram, all resistances are given in ohms and all capacities in farads except where otherwise stated.

Certain controls are required in order that the device function well. It is desirable at times to turn it off without at the same time turning off the radio, as for example when a tube fails, for in such a case the radio would be silent, or when it is desirable to increase greatly the sound volume from the speaker. Switch *AA*, which is on the front of the radio, permits *A-K* to be turned off while leaving the radio operative. It is also sometimes desirable to bring the radio back into operation before the expiration of the delay interval. This is accomplished by means of a spring switch *SS* of some kind which is also on the front of the radio, and which merely discharges capacitor *C*. The radio and *A-K* are both conveniently actuated by the same power switch.

The most important control, and the necessary one, regulates the response to sound disturbance. The device must successfully discriminate between the normal level of existing sound and the

[1] I. C. Cornog, "The electronic time-delay circuit," *Am. J. Physics* **14**, 190 (1946).

[2] W. C. Morrison, "Electronic interval timer," *Am. Photography* (April, 1948).

sharp sound representing the signal to operate. There can be extreme sensitivity, which is obviously undesirable since then the music itself would turn off the radio. Consideration of the circuit will show that there are two variables which affect the sensitivity; *viz.*, the variable resistor in the microphone circuit, and the potential divider *P* which determines the grid potential of gas triode 885. Of these the potentiometer is the more useful one for adjusting sensitivity, the resistor in the microphone circuit then being used mainly to compensate for the change of resistance of the dry cell with age.

I hold no brief for the engineering of this circuit. It just grew, like Topsy, and was used in the breadboard state until finally compressed into the usual metal cabinet, at which time the voltage doubler power supply was substituted for a conventional transformer supply. Perhaps an experienced engineer could make a better job of it, but any great simplification would be unlikely. Certainly the two voltage regulator tubes could be replaced by some sort of potential divider, and tube 2A4G would be replaced, being now outmoded.

For those interested in applying electronic circuits to various problems, *A-K* presents several possibilities. For example, if a suitable filter circuit were introduced immediately preceding triode 885, it would be possible to trigger the device by means of a sound of very small frequency range, such as a whistle of given pitch.

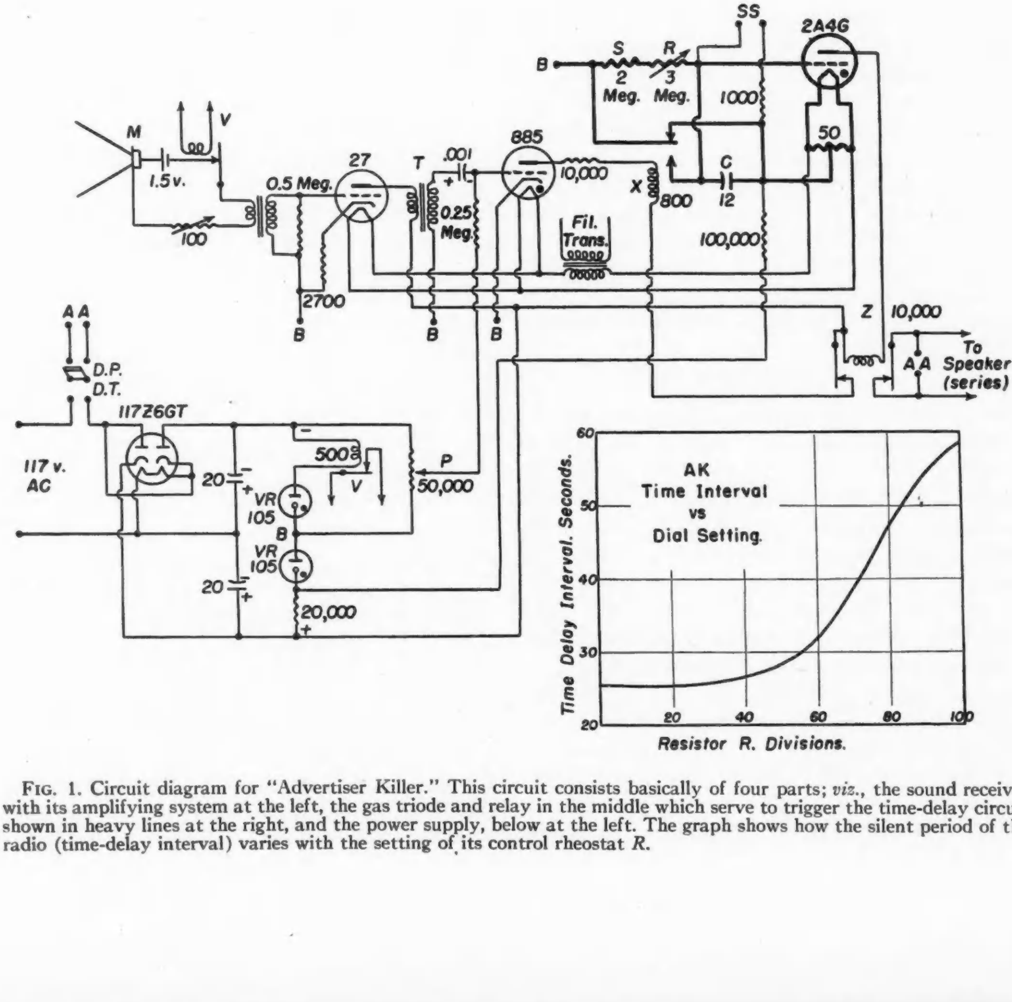

FIG. 1. Circuit diagram for "Advertiser Killer." This circuit consists basically of four parts; *viz.*, the sound receiver with its amplifying system at the left, the gas triode and relay in the middle which serve to trigger the time-delay circuit shown in heavy lines at the right, and the power supply, below at the left. The graph shows how the silent period of the radio (time-delay interval) varies with the setting of its control rheostat *R*.

There would be considerable advantage in this, for then there would be less chance of response to common sounds, or to loud music. As another example, it would be desirable to rearrange the circuit so that the device would revert to the normal state as the result of a second sound signal, say a second handclap, before the normal expiration of the delay interval.

A Laboratory Experiment on Trajectories

Floyd W. Parker
Lincoln Memorial University, Harrogate, Tennessee

THE subject of projectile motion has become well established as an important part of any course in elementary physics. It is usually met early in the course and receives a sizable fraction of the time devoted to lecture and to recitation. It is somewhat unfortunate, therefore, that a study of trajectories is generally overlooked in the laboratory. An experiment has been devised which fills this need and has the following advantages: (a) it can be presented as soon as the subject of two-dimensional projectile motion has been treated in the classroom, (b) it requires only one item of equipment not readily available in the laboratory and this item can easily be constructed, and (c) it gives the student ample practice in applying the equations associated with projectile motion to actual situations.

I. Equipment

The experiment requires a large flat surface which can be inclined at a small angle, perhaps 5° or 10° from the horizontal. The Iowa University physics laboratory, in which this experiment was initiated, had available a movable table which was approximately 4 ft wide and 15 ft long. The experiment was performed in such a way that the full length and width were utilized. The possession of such a large table, however, is not essential to the performance of the experiment and any moderately large table or other surface, such as a standard 4×8 plywood sheet can be used if proper judgment is exercised in the selection of experimental conditions.

The one special item of equipment necessary is a device which will accelerate a steel ball and send it rolling onto the table surface with a repeatable linear velocity. One such device which has proven quite satisfactory is shown in Fig. 1. It will be referred to hereafter as the launcher. It consists essentially of an inclined track down which the ball rolls to reach the table, and a plumb line to insure constancy of slope. A flexible strip of metal attached to the rails and between them, as indicated in Fig. 2, has been found helpful in preventing the initial speed of the ball from being less when projected uphill than when projected along a horizontal surface. The position of attachment of the strip is such that support of the ball is smoothly transferred from the launcher rails to the strip before the ball passes beyond the region of rigid attachment.

All other items are common equipment in any physics laboratory and need no special mention here.

II. Procedure

An outline of the procedure used in the laboratory is as follows:

1. The students were familiarized with these equations, in which the symbols are those commonly used in an elementary description of projectile motion.

$$S=v_0t+\tfrac{1}{2}at^2, \quad (1)$$
$$T=-(2v_0\sin\theta)/a, \quad (2)$$
$$R=-(v_0^2\sin 2\theta)/a, \quad (3)$$
$$H=-(v_0^2\sin^2\theta)/2a, \quad (4)$$
$$v_0=x(g/2y)^{\frac{1}{2}}. \quad (5)$$

2. The translational speed v_0 of a steel ball which had rolled down the track of the launcher and onto the surface of the table was determined by allowing the ball to drop

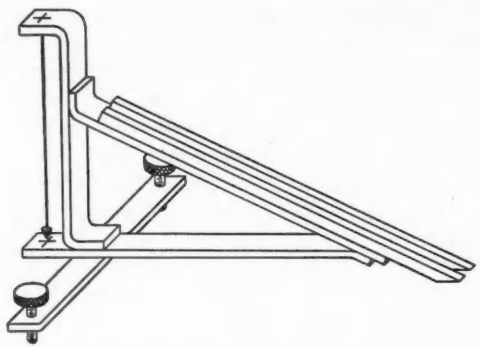

FIG. 1. Launcher used to give a steel sphere a repeatable linear velocity v_0.

to the floor as indicated in Fig. 3, recording the values of x and y, and substituting these values in Eq. (5).

3. The table was tilted as shown in Fig. 4 and the effective acceleration of the ball in the direction of the gradient was determined by projecting the ball directly up the slope as indicated at A in Fig. 4, measuring the slant height attained H, substituting this value and the known value of v_0 in Eq. (4), and solving for a, which is the acceleration of a pseudogravity acting in the plane of the table.

4. Each student selected or was assigned a particular value of θ and computed the values of T (time of flight), R (range), and H (maximum height) corresponding to that angle by use of Eqs. (2), (3), and (4).

5. The student was made responsible for the placing of suitable targets at the computed positions of R and H and for the alignment of the launcher at the proper angle. The ball was then released and was observed to traverse a path on the surface of the table similar to the path shown at B in Fig. 4, hitting the targets if the computations were correct and the conditions favorable or missing them if they were not. The computed value of T was checked by means of a stop watch.

In the elementary laboratories of the University of Iowa, at the time this experiment was first presented, each instructor had charge of approximately twelve students. The time available for the experiment was one hour and fifty minutes. Each student was given a mimeographed sheet which contained the necessary equations, a short description of the methods used in determining v_0 and a, together with appropriate sketches and a table in which the values of T, R, and H, could be inserted corresponding to values of θ of 15°, 30°, 45°, 60°, and 75°. In addition, further information concerning an auxiliary experiment, which will be described later, was included.

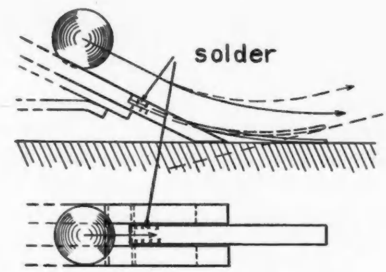

FIG. 2. Modification of launcher which prevents initial speed v_0 of a ball from being less when projected uphill (dashed lines) than when projected along a horizontal surface.

The first ten or fifteen minutes of the time available was used to outline the experiment to the students and to re-acquaint them with the formulas to be used. Equation (2) was derived by use of Eq. (1) with $S=0$ and with v_0 replaced by $v_0 \sin\theta$. Equation (3) followed immediately by multiplying each side of Eq. (2) by $v_0 \cos\theta$. Equation (4) was obtained by combining Eqs. (1) and (2), with v_0 of Eq. (1) replaced by $v_0 \sin\theta$; and t of Eq. (1) replaced by $T/2$ from Eq. (2). Equation (5) was derived by use of Eq. (1) with reference to a sketch similar to Fig. 3.

The instructor then demonstrated the technique to be used in the determination of v_0 by placing the launcher near the end of the table, adjusting the leveling screws until the plumb line was centered, and carefully releasing a steel ball at the top of the incline. After rolling down the incline, the ball would traverse the short distance to the table edge and fall onto the floor. This is the situation indicated in Fig. 3. Carbon paper was utilized in determining the position of the striking point. A number of repeat runs were then made during which student participation was encouraged to the extent that

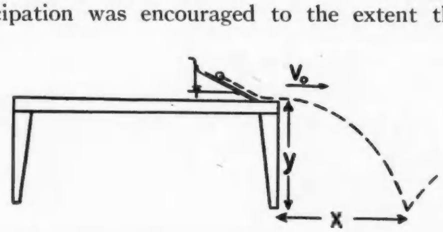

FIG. 3. Illustration showing the arrangement utilized in the determination of v_0.

such functions as operating the launcher and adjusting the carbon paper at the striking point were being performed by students.

The instructor then divided the class into two groups and assigned to one the task of measuring the appropriate distances and calculating the value of v_0. The other group assisted the instructor in determining the value of a. The table was tilted with one edge above the other as shown in Fig. 4 and the launcher was mounted on an adjustable laboratory stool or small table in such a way as to project the ball directly up the incline with the initial velocity v_0 as indicated at A of the figure. The slant height H attained by the ball was recorded.

By this time the first group was ready with the computed value of v_0 and the second group, using Eq. (4), quickly calculated the value of a. Since the initial velocity enters as v_0^2 in Eqs. (3) and (4), v_0^2 as well as v_0 and a was recorded by each student for use in future computations. The time elapsed at this point was approximately 40 minutes.

The class was then redivided into groups of three or four students and each group computed T, R, and H for the particular angle assigned to it. After the computations were completed, each group was given an opportunity to test the correctness of its predictions. The launcher and the adjustable stool or table which supported it were placed at the corner of the large table and the launcher was aligned at the correct angle by the members of the group. A small matchbox was placed at the predicted range position and a small cork stopper at the predicted position of maximum height as indicated at B in Fig. 4,

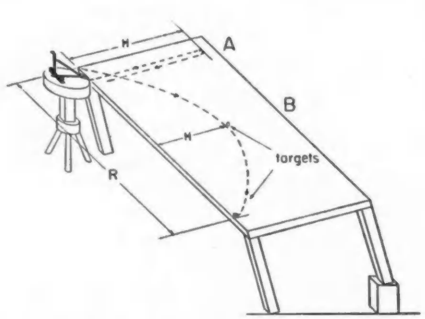

FIG. 4. Illustration showing the arrangement utilized in the determination of a and in the check shots.

all decisions and adjustments being made by the group concerned. When the targets had been placed and all was in readiness, the shot was made. No practice shots were allowed. The total elapsed time at the completion of these shots was approximately 90 minutes.

It is well to point out that the 90 minutes allowed for the completion of the experimental work described above applies only if the experiment is well organized and supervised by a competent instructor. Students working by themselves with only occasional direction from an instructor could not be expected to complete the outlined experimental work in less than two and one-half to three hours.

III. Discussion

In the presentation of this experiment to students taking introductory physics many considerations were ignored which would have received attention in a class of more advanced character. The effects of friction and rotation, for example, were not mentioned unless they were first brought into the discussion by students, since it was felt that the danger of losing sight of the objective of the experiment because of an overabundance of subsidiary discussions was greater than the dangers associated with the avoidance of such discussion. Under other circumstances these effects cannot readily be ignored and should be discussed fully.

In the initial conception of the experiment, provision was made to compensate for rolling friction by adjusting the slope of the table until no measureable change in speed occurred as the ball traveled the length of the table. This condition was known to be achieved when the value of x (Fig. 3) could be shown to be independent of the length of table traversed by the ball. The adjustment thus made, however, is correct only if the ball is moving parallel to the long dimension of the table and in the proper direction. For motion parallel to the ends of the table surface, the adjustment is quite incorrect. Consequently, compensation for friction in the manner described above is not recommended.

The effect of rotation, unlike that of friction, is not small. This complication has been avoided, however, by choosing a method of determining

the value of a which does not involve the moment of inertia of the ball. It is well known that the ratio of the linear accelerations of an ideal rolling sphere and an ideal sliding particle under the same conditions is 5/7. Consequently, a rolling sphere, and a sliding particle on which forces have been reduced by this ratio will have the same acceleration and therefore the same trajectory. Since the trajectory of the sliding particle is known to be parabolic, that of the sphere must be parabolic also, and the use of the rolling ball analog in the study of other types of parabolic trajectories is thus justified.

IV. Results

The primary objective of the experiment, as performed and as described above, is neither the attainment of high precision nor the complete understanding of the rolling ball analog, but rather merely to give the students an opportunity to apply the equations of projectile motion to concrete situations. This being the case, it is only necessary to ensure that the precision, judged by the percentage of hits, is not so low that it arouses derisive reactions from the students. The percentage of hits scored will depend, among other things, on (a) the size of the ball and of the targets, (b) the ability to align the launcher properly at the correct angle, and (c) on the condition of the table top or other surface used. The launcher design of Fig. 1 is not without shortcomings. Although the launching of a ball along a given line on a horizontal surface is quite simple, a similar launching along a line *on a tilted surface* is accomplished with somewhat less ease and precision. Likewise, the table surface used by the author was far from ideal. A certain amount of sag and warping was easily detected by sighting along its surface. In spite of these imperfections, however, the probable error in the values of v_0 and a obtained by 24 separate laboratory sections was not greater than two percent. The steel balls used were $\frac{3}{4}$ in. in diameter. The small cork stopper used to mark the position of maximum height had a diameter of approximately $\frac{1}{4}$ in. and the matchbox target placed at the range position had a dimension in the plane of the table and normal to the trajectory of approximately 2 in. With these targets placed at the computed positions, the percentage of hits for the 24 laboratory sections was greater than 50 percent. Judging from the comments of a large number of students, this percentage was not considered by them to be low and the general reaction to the experiment was extremely favorable.

V. Additional Experiment

It has been mentioned above that with efficient utilization of time the experiment could be completed in 90 min. of the 110 min. available. Consequently, it was possible to perform an additional experiment which used the remaining time and extended the scope of the work performed during the laboratory period to include a problem involving the flight of a projectile acted upon by gravity. The experiment is essentially one in which the students predict the striking point of a metal ball which has been launched through an open window situated considerably above ground level. It should be made clear that this additional experiment is by no means an essential part of the experiment discussed above. It can be deleted if the time available is insufficient, or for other reasons. However, if the laboratory period is sufficiently long, and if proper facilities are available, this additional experiment is quite worthwhile.

A spring gun, such as is used in ballistic pendulum experiments, was clamped securely to a small laboratory table which was properly placed and tilted in order to allow the gun to fire a metal ball through an open window, the angle of elevation being 45°. The situation is shown schematically in Fig. 5. A similar sketch was included on the mimeographed sheets presented to the students at the beginning of the laboratory period.

In order to predict the position at which the projectile would strike the ground, it was first necessary to know the initial velocity v_0 imparted to the projectile launched at an angle of 45° by the gun. It is obvious that this value will differ from the value of v_0 for a horizontally projected ball. The proper value was determined by moving the table supporting the gun so that the range of the projectile, launched at 45°, could be measured within the room. An adjust-

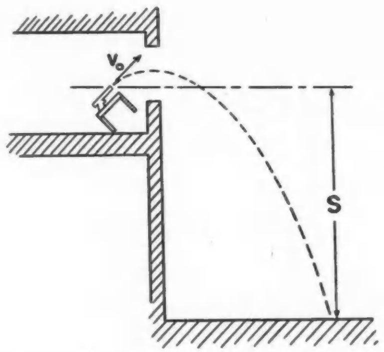

FIG. 5. Drawing showing the essential features of an additional laboratory experiment involving the prediction of the striking point of a projectile.

able stool was placed so that its surface was at the same height as the ball at the instant it left the gun and at the proper distance to intercept its trajectory. The measured value of the range was then inserted into Eq. (3) to obtain v_0.

The table and gun were then moved to the window and placed in a predetermined position. The horizontal and vertical distances from the muzzle of the gun to a reference point on the ground were given to the students. Using the given vertical distance S and the vertical component of v_0 in Eq. (1), the time of fall was determined. The product of the horizontal component of v_0 and the time of fall gave the predicted horizontal distance to the striking point. A colored paper target, such as can be obtained wherever archery equipment is sold, was placed at the computed striking point by the students. The direction of flight was ascertained by firing another metal ball whose mass differed appreciably from that of the ball under consideration.

When all was in readiness and the students were clustered at the windows or near the target, the trigger of the gun was pulled, the projectile was launched through the window and fell to the lawn below striking in or near the bulls-eye.

An indication of the accuracy which can be expected can be gained from the results achieved by 24 laboratory sections. The value of v_0 was found to be approximately 15 ft sec^{-1} and the vertical distance to the striking point was approximately 50 ft. Under these conditions 12 of the 24 shots attempted hit within 5 in. of the computed striking point.

In recent months a 10-in. diameter gong has been substituted for the paper target with considerably increased effectiveness when judged by audio standards.

The author wishes to thank Dr. C. J. Lapp for his helpful counsel and Mr. J. G. Sentinella for his assistance in constructing the launcher.

Prior to this century, an analysis of the experimentalist's activity might have shown that the bulk of his time was spent in getting ideas and in analyzing the data of his subsequent experiments, while a minimum of time was spent in the construction of instruments. In the present period, too often the scientific situation is such that the bulk of his time has to be spent in the construction of instruments. From the point of view of research this is bad. . . .—E. U. CONDON, Is There a Science of Instrumentation? (*Science* **110**, 339 (Oct. 7, 1949)).

The Teaching of Electricity and Magnetism at the College Level*

II. Two Outlines for Teachers

(Report of the Coulomb's Law Committee† of the A. A. P. T.)

3. Acceptable Schemes for Teaching the Fundamental Principles of Electricity and Magnetism

3.1 Introduction

IN the course of the committee deliberations several possible acceptable procedures for teaching the fundamental principles of electricity and magnetism were discussed. Two of these outlines were worked out rather fully and are reproduced herewith as Part 3 of this report. They represent the best thought of the committee, but lay no claims to inspiration. Numerous acceptable variations from each of the basic outlines will suggest themselves to the experienced teacher.

For the sake of clarity the outlines are given in considerable detail. Material included in small print for completeness is considered suitable for an intermediate course in electricity and magnetism rather than for an introductory course in general physics. In working out the basic outlines (large print) we have had in mind primarily the needs of first-year courses for students intending to continue into engineering, physics, or chemistry. The selection of the more descriptive material usually given in the freshman year to liberal arts students who do not expect to major in physical science can be made in so many ways that it seemed unwise to attempt it here. We believe, however, that the teachers of such students have the same responsibility for fostering clear thinking as teachers of engineering and physics majors. It is our hope that the outlines here given will be helpful to both classes of teachers even if many find it necessary to omit some of the logical connections.

3.2 Outline I. *Traditional Approach*

3.21 The Electrostatic Field in Free Space.—In order to gain the immediate interest of practically minded students who have an initial familiarity with the elementary phenomena of current electricity it will be well to point out at the beginning of electrostatics that it is not possible to acquire clear ideas about current electricity without a knowledge of the elementary facts of electrostatics. One can also call attention to the practical importance of the Van de Graaff generator, the electrification of powders, and atmospheric electricity.

Give a qualitative demonstration of the elementary phenomena of "frictional" electricity during which the existence of attractive and repulsive forces is made clear. Use the Faraday ice-pail experiment to show that the presence of an electric charge on a body may be detected by the deflection of an electroscope. This basic operation for the detection of charge should be demonstrated in a lecture experiment. Define *equal charges* as those that produce equal electroscope deflections. Show that two such equal charges introduced simultaneously into the ice-pail produce either zero deflection, or a larger deflection than does either charge alone. Use this demonstration and another in which contact charges are generated inside the ice-pail to justify the "plus" and "minus" conventions for charges.

Explain how to calibrate the electroscope in terms of an arbitrarily chosen standard charge and integral multiples of that charge. Point out that the calibrated electroscope can be used to measure the ratio of two charges without regard to the ultimate unit of charge to be adopted.

Develop Coulomb's law in the forms of Eqs. (1–1) and (1–2) as an experimental fact, employing previously measured charges, and proceed to the specification of the constant of proportionality k_0 in the system or systems of units to be employed. In this development of Coulomb's law take due account of the logical steps (*c*), (*d*), (*f*), (*j*), (*l*), (*m*), and (*n*) at the end of Part 1.[38]

* Continued from *Am. J. Physics* **18**, 1 (1950). The list of principal symbols is repeated here.

† The Committee is W. F. Brown, Jr.; N. H. Frank; E. C. Kemble, *Chairman;* W. H. Michener; C. C. Murdock; D. L. Webster.

[38] This procedure based on the Faraday ice-pail experiment is much less involved than a logical development

A still more simplified alternative approach permissible in an introductory course is to adopt a frankly quasi-deductive or postulational point of view after introducing the inverse square law in the form given in (*c*), and the definition of a point charge as in (*d*), Sec. 1.2. One can then say that a detailed study of the total experimental situation warrants the postulate that the algebraic value of a point charge can be so defined that the complete relation between force, charges, and distance for point charges becomes Eq. (1–2). It follows from this postulate that a unit charge must be one which repels an equal charge at distance r with a force k_0. The algebraic charge q on any body must be equal to Fr^2/k_0, where F is the algebraic force of interaction with a unit positive charge at a distance r large enough to make the inverse square law applicable. Thus the definition of charge is deduced from the postulate.[39]

Observe that according to Coulomb's law the force on a small[40] test-charge at a point P in an electrostatic field can be resolved into the product of two factors, one of which is the algebraic value of the test-charge itself, while the other is a vector quantity $\mathbf{E}$ depending on the location of other nearby charges, but independent of both the magnitude and the sign of the test-charge. Designate the factor $\mathbf{E}$ as the *electric intensity* at the point P.[41] Note that the total electrical intensity due to the combined influence of a distribution of point charges is the vector sum of the electric intensities which the several point charges would produce if each acted alone.

Discuss the general concept of a "field." Explain the meaning of the term "line of force" in an electrostatic field and exhibit such lines of force by the use of gypsum crystals or cork filings. Call attention to the fact that the lines of force converge as we pass to points of the field where the electric intensity is strong and diverge as we pass to points where it is weak. Explain also the use of representative lines of force (representing unit tubes of force) to indicate graphically the direction and magnitude of the electric intensity in the Coulomb field of a point charge. Make it clear that such use of representative lines of force is possible only because the field of a point charge follows an inverse square law. State the more general principle that unit lines of force can be used in the same way to represent a general electrostatic field in charge-free space. Formulate the *theorem of Gauss*, using the number of unit lines of force which emanate from a closed surface as a representation of the flux of the vector $\mathbf{E}$ through the surface. Apply the theorem to show that in the absence of electric currents no charge can exist in the interior of a conducting medium. Demonstrate with gypsum crystals or cork filings the rule that the electric intensity in space just outside the surface of a conductor in static equilibrium must be normal to the surface. Apply the theorem of Gauss to show that the algebraic value of the normal component of $\mathbf{E}$ in free space next to such a conducting surface must have the numerical value

(Gauss) $$E_n = 4\pi\sigma_e, \tag{3–1a}$$
(Giorgi) $$E_n = \sigma_e/\epsilon_0. \tag{3–1b}$$

Define *electric potential* as work-per-unit-charge done by or against electrostatic forces, emphasizing the fact that according to Coulomb's law the work of moving a test charge from one point to another in an electrostatic field is independent of the path. State, or if time permits, evaluate, the potential distribution in the neighborhood of an isolated point charge or a distribution of such charges. Define the *volt* as a potential difference of one joule per coulomb.

Define the *capacitance* of a conductor and of a condenser. Develop the theory of the infinite parallel-plate air condenser from Eq. (3–1). Evaluate the energy of a charged condenser.

Discuss Millikan's oil-drop experiment and the atomicity of electric charge. Note the possibility of interpreting the charge on any body in terms of a counting operation.

based solely on the Coulomb's law experiment such as that attributed to the "ultra-conscientious" instructor in Part 1.

[39] This second procedure has an obvious advantage in simplicity and brevity. Its chief disadvantage is that it fails to separate the elements of the theory which are conventions from those which are required by the experimental facts and thus tends to familiarize the student with the formalism of the theory without giving him a firm grasp of the experimental significance of the concepts.

[40] See Eq. (2–14).

[41] It may be helpful to point out that the "acceleration of gravity" g is the magnitude of the force-per-unit-mass exerted by the earth's gravitational field in a small body. The vector $\mathbf{g}$ is accordingly the gravitational analog of the electric intensity $\mathbf{E}$.

If the time allowance and the mathematical level of the course permit, prove the theorem of Gauss by a discussion of the surface integral and show that in charge-free space the flux of **E** through a tube of force is constant. On this basis validate the quantitative representation of the electric intensity by unit lines of force.

3.22 The Electrostatic Field in Dielectric Media.—This subject may be introduced along the lines indicated by paragraphs in small type, Sec. 2.4, stressing the electric polarization as the basic concept of the theory.

In applying the theory to the infinite parallel-plate condenser filled with an ideal homogeneous isotropic dielectric assume that the electric polarization **P** is given by

(Gauss) $$\mathbf{P}=\chi_e\mathbf{E}, \tag{3–2a}$$

or (Giorgi) $$\mathbf{P}=\epsilon_0\chi_e\mathbf{E}, \tag{3–2b}$$

where the electric susceptibility χ_e is positive and depends only on the nature and state of the medium, being independent of **E**. Note that Eq. (3–2) is not in harmony with the experimental behavior of *all* insulating media.

Assume from the symmetry of the problem that **E** and **P** are normal to the plates throughout the space between them. Then apply Gauss's theorem to prove that each of them has the same value at all points of the field. Postulate, in accordance with the definition of **E** laid down in Sec. 2.4, that Eq. (3–1) is to be generalized to become

(Gauss) $$E_n=4\pi(\sigma_c+\sigma_p), \tag{3–3a}$$

or (Giorgi) $$E_n=(\sigma_c+\sigma_p)/\epsilon_0, \tag{3–3b}$$

where the polarization charge σ_p is equal to the component of **P** directed from the dielectric toward the plate which carries the conduction charge σ_c. Hence, show that

(Gauss) $$E_n=4\pi(\sigma_c-P_n)=4\pi(\sigma_c-\chi_eE_n), \tag{3–4a}$$

or (Giorgi)

$$E_n=(\sigma_c-P_n)/\epsilon_0=(\sigma_c/\epsilon_0)-\chi_eE_n. \tag{3–4b}$$

Finally solve for σ_c/E_n and work out the formula for the capacitance in terms of the relative dielectric constant κ_e, defined as $1+4\pi\chi_e$ in the Gaussian unit system, or $1+\chi_e$ in rationalized Giorgian units.

If time and the course level permit, one may proceed to the discussion of the field due to a point charge q embedded in an infinite ideal homogeneous and isotropic dielectric. Derive Coulomb's law in the form $E=q_c/\kappa_e r^2$ (Gauss), or $E=q_c/(4\pi\epsilon_0\kappa_e r^2)$ (Giorgi) from the theory of polarization.[42] Note that the equation holds if we have a number of point charges, or a distribution of charge on conductors imbedded in an infinite homogeneous dielectric. Call attention to the fact that it fails if the dielectric does not fill all the space in which there is a field, or if the dielectric is inhomogeneous.

Show by the method described at the end of Sec. 2.1, that the net force on a free point charge q_c in a dielectric fluid must be $q_c\mathbf{E}$. Thus the same experimental method can be used to measure **E** in a fluid as in a vacuum.

Discuss the significance of the electric intensity **E** and the potential V in the interior of a solid dielectric. From the point of view of the continuum model, **E** is a vector quantity calculated according to Coulomb's law for a vacuum from the combined distribution of conduction charges and polarization charges. The relation between **E** and the potential gradient is the same inside the dielectric as in a vacuum. If time permits, introduce **D** as $\mathbf{E}+4\pi\mathbf{P}$, or as $\epsilon_0\mathbf{E}+\mathbf{P}$, and show that the flux of $\mathbf{D}/4\pi$, or **D**, from any closed surface is equal to the total conduction charge inside the surface. Thus the vector lines of **D** terminate only on conduction charges.

If time permits, mention also the Kelvin-cavity definitions of **E** and **D** for the interior of a solid dielectric (see Sec. 2.4, small type).

It is desirable that the polarization vector be discussed from the atomic point of view as well as in terms of the continuous model. The application of an external field may polarize an atom or molecule, or may change the orientation of its pre-existing dipole moment. Show that a material composed of polarizable atoms and molecules

[42] For an elementary derivation one may start with the fact that symmetry demands that the polarization charge-density must be uniform on any sphere drawn about the point charge q_c as a center. Furthermore, **E** and **P** must be everywhere radial. Show by the theorem of Gauss that the radial component of **E** is

$$(q_c+q_p)/r^2 \quad \text{or} \quad (q_c+q_p)/(4\pi\epsilon_0r^2)$$

depending on the units employed, where q_p is the total polarization charge inside a sphere of radius r centered on the point conduction charge q_c. If $P(r)$ and $E(r)$ denote the radial components of **P** and **E** on the sphere,

(Gauss) $$q_p=-4\pi r^2P(r)=-4\pi r^2\chi_eE(r), \tag{a}$$

or (Giorgi) $$q_p=-4\pi r^2\chi_eE(r)\epsilon_0. \tag{b}$$

Combining this equation with the previous expression for $E(r)$ show that q_p is independent of r and must therefore adhere directly to the point conduction charge q_c. Thus the polarization charge density is zero except at the center of the sphere. (In an advanced course one would prove this from divergence relations.) Elimination of q_p from the two equations which relate it to E gives Coulomb's law.

can be expected to behave like the previously introduced continuous model of a dielectric. Identify the polarization **P** with the product of the number of molecules per unit volume and the average vector dipole moment of the molecules. Note that the vector **E** of macroscopic electrical theory is a smoothed-out vector function obtainable in principle by averaging the irregular microscopic field over a volume large enough to contain many molecules.

3.23 The Magnetostatic Field in Free Space.—Begin by exhibiting the basic qualitative facts of the magnetostatics of permanent magnets in air, emphasizing the North-South orientation of a freely suspended needle-magnet and the existence of small regions (poles) of intense apparent magnetic activity of different sign which appear at opposite extremities of a properly magnetized slender rod of iron or steel. Observe that magnetostatic experiments in air, like electrostatic experiments are, to all intents and purposes, equivalent to similar experiments performed in a vacuum. Describe the Coulomb experiment and postulate the existence of positive and negative distributions of magnetic pole strength, or magnetic charge, on the surface and within the volume of magnets, these poles having the property of exerting forces on one another which may be computed from Coulomb's law in the form

(Gauss) $F_m = (pp')/r^2$, (3–5a)

(Giorgi-Kennelly)[43] $F_m = (pp')/(4\pi\mu_0 r^2)$. (3–5b)

Point out that on the basis of this assumption it is possible to develop a theory of magnetostatics analogous to that which has just been developed for electrostatics. In the process of this development there will appear an intensity of the magnetic field **H** analogous to **E**; a magnetic polarization **M** analogous to **P**; a magnetic susceptibility χ_m analogous to χ_e; a magnetic induction **B** analogous to **D**; a dimensional constant μ_0 analogous to ϵ_0; and a relative permeability κ_m analogous to κ_e. Complete the implicit definition of pole-strength in the Giorgi-Kennelly system by assigning the value $4\pi\times10^{-7}$ to μ_0, deferring the justification of the choice till later. Define the magnetic intensity **H** in free space as force-per-unit-pole-strength on a needle-magnet pole.

Deduce the existence of the earth's magnetic field from the compass experiment. Perform the floating magnet experiment and deduce the rule that the algebraic sum of the pole-strengths of a magnet is zero. Perform the experiment of breaking a magnet in two and observe that thereby one does not separate positive and negative poles, but produces two complete magnets. Hence we may infer that permanent magnets are the seat of a permanent distribution of magnetic polarization **M** similar to the electric polarization **P** observed in dielectrics (Sec. 3.22) except that it is not caused by an external field. On this hypothesis the distributions of magnetic poles at the ends of a bar magnet are similar to the distributions of bound electric charge on the surfaces of a polarized dielectric. The hypothesis is in satisfactory agreement with the outcome of the Coulomb experiment, both for large distances and for small distances.

Here it may be noted that many low symmetry crystals are by their structure permanent electrets, but that the electret properties are not commonly observed because of the neutralizing charges which cover the surface of the crystal due to the conductivity of the air. The fact that permanent magnets are not similarly neutralized may then be cited as evidence that no free poles exist in the atmosphere, i.e., that air is not a magnetic conductor. It should be pointed out that we do not find anywhere in nature the magnetic analog of the electric conductor nor free poles analogous to free electric charges. Since the only magnetic poles known to exist are those due to the magnetic polarization of matter, it becomes evident that the elementary particle of magnetostatics is not the magnetic pole, but the magnetic dipole.

Demonstrate the measurement of the magnetic moment of a magnet and of the strength of the earth's field by a magnetic pendulum and magnetometer.

Discuss the elementary facts regarding induced magnetism. Give a qualitative explanation of the orientation of iron filings in a magnetic field in terms of induced magnetic moments. Interpret magnetic polarization as a distribution of elementary magnets with predominant ori-

[43] Since the Kennelly definitions of pole-strength, magnetic moment, and magnetization are better adapted to the traditional approach to magnetism described in this outline (Outline I), they are used throughout Sec. 3.2. To get the corresponding equations in the Sommerfeld-Stratton scheme replace p by $\mu_0 p$, and **M** by μ_0**M**. Cf. footnote 33, Sec. 2.6.

entation in the direction of **M**. Call attention to the fact that an electric circuit produces a magnetic field like that of a permanent magnet and explain that there is good reason to identify elementary magnets with molecular, or atomic, currents of electricity. From this point of view the poles at the ends of a magnet are convenient fictions, the forces and torques acting on magnets being due in the last analysis to interactions between the external magnetic field and the molecular currents. Develop the qualitative molecular theory of magnetism in the usual way, calling attention to the Curie point and allied phenomena.

3.24 Electric Currents.—Begin with the distinction between ***conduction currents*** and ***convection currents***. If the instructor so desires he may mention also the existence of ***polarization currents*** and so-called "displacement currents" involving no transport of charge, but producing magnetic effects similar to those of a true flow of charge.

Show the transfer of charge which occurs as a result of the transient conduction current that takes place when insulated conductors at different initial potentials are connected by a wire. Explain that *steady* conduction currents require closed circuits and nonelectrostatic forces originating in an electric "pump," such as a battery or a dynamo.

Exhibit a steady current and show qualitatively the heating effects, magnetic effects, and chemical effects by which currents can be detected and measured. Define the emf of an electric pump as the work-per-unit-charge done by that pump when current passes through it. Emphasize that the emf is not really a force in the modern sense of that term, being measured in volts, or joules-per-coulomb, and not in dynes, or even in dynes-per-coulomb.[44]

[44] Note that the net force in dynes acting on an element of conduction charge in a solid or liquid conductor is not practically important because the frictional resistance to motion is so great that the acceleration and kinetic energy of the charges are always extremely small. For this and other reasons most problems connected with current flow in linear circuits can be handled most expeditiously by energy methods in which emf and potential are of central importance. In the equations that govern the flow of currents in such circuits the emf plays a role analogous to that of the external force in the familiar equation for the drag on a body moving with constant speed through a viscous medium. Hence the term "electromotive force" makes sense functionally in spite of the fact that it is misleading from the point of view of dimensions.

Call attention to the fact that, if a battery is on open circuit, its terminals are charged to a difference of potential equal to the emf. Current flow through the cell is prevented by the balancing of the nonelectrostatic thermochemical forces in one direction against the opposing electrostatic forces due to the charge distribution on the terminals and inside the battery. Explain that when the circuit is closed, so that the current can flow through the battery and the external wire without building up additional charges on the terminals of the battery, ***the electrostatic field due to the battery is radically altered***. The potential difference between the terminals is lowered and a steady fall of potential as one passes along the wire from the positive terminal of the battery to the negative is established. The fall of potential as one proceeds along the wire can be measured with a sensitive electroscope. It is independent of the shape of the wire, whether wound in a coil or straightened out. Observe that the redistribution of potential in the space around the battery is due to the creation of a distribution of charge on the surface of the wire controlled primarily by the frictional resistance which the current meets in the wire.

Define the current I in a conducting channel conceptually as the number of units of charge crossing a section of the wire per second. The flow of positive charges in one direction is equivalent to an equal flow of negative charges in the opposite direction, so that the net current is actually the sum of two components due to the two types of charge. Note that in solid conductors the flow of positive charges is insignificant, but in liquid and gaseous conductors this is not true.

Remind the students that electric currents produce magnetic fields and that through these fields they exert forces and torques on each other. Enunciate, or demonstrate, the law that the force (or torque) exerted by the simultaneous influence of two different current-carrying coils A and B on a permanent magnet or on a third current-carrying coil in free space is the vector sum of the forces (or torques) that A and B exert when acting separately.[45] Draw the inference that the force (or torque) that a current-

[45] A demonstration of this law using a current balance or a tangent galvanometer with two large coils might well be made the basis of a laboratory experiment.

carrying coil A exerts on a magnet, or on a second coil, must be proportional to the current in A. Define the electromagnetic unit of current and/or the ampere in terms of the magnetic intensity at the center of a circular coil or in terms of the force exerted by a current in a very long straight wire on a parallel and equal current. Define the coulomb in terms of the ampere. If you use cgs units, introduce the ratio c of the abampere to the electrostatic unit of current.

Joule's law—that in a given homogeneous linear conductor at fixed temperature and pressure the rate at which energy is dissipated in heat is proportional to the square of the current—can be introduced to advantage before Ohm's law. *Resistance* can then be defined as the ratio of the power dissipated to the square of the current. Develop the law for the addition of resistances. Note that the resistance of a wire is inversely proportional to its cross-sectional area and define specific resistance.

Ohm's law for a complete circuit can then be deduced from Joule's law as a simple application of the principle of the conservation of energy. Note the failure of both laws for currents in gases. Explain how to evaluate the total electromotive force when the circuit contains several batteries or generators in series. Formulate *Ohm's law for a part of a circuit* consisting of a pure resistance, or of a resistance and battery. Show that this law, verifiable by electrostatic measurement of the potential drop, in conjunction with Joule's law, implies that the work done by electrostatic forces on the current in each section of an electric circuit is equivalent to the heat developed in that section. Discuss the changes in the potential difference between the terminals of a battery due to the flow of current through its internal resistance.

At this point it will be advisable to discuss a variety of applications of Ohm's law, including the use of galvanometers, ammeters, and current voltmeters. In this connection the measurement of the capacitance of a condenser with a ballistic galvanometer, or with an alternating current ammeter and voltmeter should be described (see footnote 49, Sec. 3.32). Then call attention to the fact that if such a measurement is carried out with an "air" condenser whose capacitance is geometrically calculable, the observations yield a determination of ϵ_0 or c, according as one is using Giorgi units or Gaussian units.

Returning to the magnetic fields of steady currents in linear circuits, demonstrate the qualitative similarity between the external fields of a current-carrying coil and of a permanent magnet at distances large compared with the linear dimensions of the coil and magnet. On the basis of experimental data postulate the quantitative equivalence of the fields of a plane current loop and a magnet of suitable moment at large distances from the sources. State, or derive, Ampere's theorem of the equivalence at all distances of the field of a linear circuit carrying a current I and the field of a magnetic shell whose moment per unit area is proportional to I and whose periphery follows the line of the circuit. Note that the required moment per unit area is equal to I in the absolute electromagnetic and Sommerfeld units, but to $\mu_0 I$ in the Kennelly mks units.

If the level of the course permits, formulate Ampere's law for the production of a magnetic field by a current in a circuit of any shape.

Call attention to the fact that in accordance with the law of action and reaction the force and torque acting on a current loop in free space due to an external magnetic field must be equal to the force and torque on an equivalent magnetic shell in the same field. Derive the law for the force on a current element in a transverse magnetic field, perhaps by considering the interaction of a circular circuit with a magnetic pole at its center.

If mks units are used, it will be well to define **B** for points in free space in terms of the force-per-unit-length, per-unit-current, in a transverse magnetic field before deriving the corresponding expression in terms of **H**. The value of μ_0 can then be referred to an experimental determination of the ratio B/H carried out by observing the torque exerted by a large circular circuit on a small circuit at its center when the axes of the circuits are at right angles.

Electromagnetic induction in free space may be introduced at this point, care being taken to distinguish between its three different forms. Explain motional emf first as a consequence of the force exerted by a magnetic field on the convection current of conduction electrons created

by displacing a conductor in the field. The magnitude of the motional emf can be worked out on this basis. Exhibit the corresponding emf due to the displacement of the sources of a magnetic field relative to a stationary coil and note that according to the theory of relativity the emf due to the relative motion of source and coil cannot depend on the absolute motion of either one. Finally, point out that a change in the flux through a stationary coil due to an alteration in the current in a neighboring coil produces the same emf as that obtained when a similar change in flux occurs as a result of relative motion.[46] Formulate the general expression for the total emf in a circuit

$$\text{(Gauss)} \qquad \mathcal{E} = -\frac{N}{c}\frac{d\Phi}{dt}, \qquad \text{(3–6a)}$$

$$\text{(Giorgi)} \qquad \mathcal{E} = -N\frac{d\Phi}{dt}. \qquad \text{(3–6b)}$$

Note that this combined expression for the emf due to three superficially distinct causes does not hold if the current is not confined to a definite linear circuit (for example, in Barlow's wheel or any system with sliding contacts) and does not indicate the parts of the circuit in which the emf is localized. It will be well at this point to drive in the distinction between emf and potential difference (commonly lumped together as "voltage") by calling attention to the case of a symmetrical ring in a decreasing axially symmetric magnetic field, where current flows through a resistance, but where it is clear from symmetry considerations that no potential difference can exist.

Call attention to the fact that the evaluation of the flux in a magnetic medium requires further discussion.

3.25 Magnetic Materials.—The discussion of magnetic materials may well begin with a more detailed interpretation of magnetization in terms of Amperian currents. Return to the qualitative polarization theory of Sec. 3.23 and note that the magnetic polarization vector **M** can be defined either as magnetic moment per unit volume, or as a vector quantity whose magnitude is the pole-strength per unit area of a transverse section. Show that an ideal uniformly magnetized cylinder would be equivalent externally to a pile of magnetic shells of equal strength and thickness and hence would produce an external field identical with that of uniform distributions of positive and negative pole-strength across the end surfaces. Show that since a shell of polarization **M** and thickness dl is equivalent to a peripheral current of magnitude Mdl (electromagnetic or Giorgi-Sommerfeld units), or Mdl/μ_0 (Giorgi-Kennelly units), the complete magnet must produce an external field **H** in free space equal to that of a uniformly wound solenoid with n turns per unit length and a current of M/n units in the electromagnetic or Giorgi-Sommerfeld systems, but of $M/n\mu_0$ amperes in the Giorgi-Kennelly system. The accepted Amperian current interpretation of magnetization implies the existence of a net uncancelled Amperian surface current whose linear density is M (or M/μ_0). (The average vector volume density of uncancelled Amperian current in any macroscopic element in the interior of the medium is proportional to curl**M** and vanishes when the magnetization is uniform.) Thus the net current distributions for an ideal uniformly polarized cylindrical magnet and an equivalent solenoid are effectively identical except that in one case we have to do with uncancelled Amperian current and in the other case with directly measurable conduction current. Point out that real cylindrical magnets are *not* uniformly polarized and that their poles are *not* distributed exclusively over the end faces.

State, or prove, that it is possible to generalize these conclusions and to derive mathematically equivalent expressions for the external field produced by a space distribution of magnetization in terms of the corresponding distribution of pole strength or of uncancelled Amperian currents.

In more advanced courses the following equations (in Giorgi-Kennelly units) may be derived.[47]

[46] Cf. L. Page and N. I. Adams, *Am. Physics Teacher*, **3**, 51 (1935), for an illuminating discussion of the logical hazards in teaching the laws of induction.

[47] The corresponding equations in Gaussian units may be obtained by replacing $\mathbf{B}/\mu_0$ by **B** and omitting the factors $1/4\pi\mu_0$ outside the integral signs. To convert to Giorgi-Sommerfeld units replace **M** by $\mu_0\mathbf{M}$. In the second and third integrals contributions from the bounding surface of the magnet have been omitted on the assumption that they are to be subsumed into the volume integrals by means of the usual transition-layer device. The vector $\mathbf{1}_r$ is, of course, the unit vector from the integration point to the field point.

$$\mathbf{H}=\mathbf{B}/\mu_0=\frac{1}{4\pi\mu_0}\iiint\frac{[-\mathbf{M}+3(\mathbf{M}\cdot\mathbf{1}_r)\mathbf{1}_r]}{r^3}d\tau, \quad (3\text{–}7)$$

$$=-\frac{1}{4\pi\mu_0}\iiint\frac{\mathrm{div}\mathbf{M}}{r^2}\mathbf{1}_r d\tau, \quad (3\text{–}8)$$

$$=\frac{1}{4\pi\mu_0}\iiint\frac{[\mathrm{curl}\mathbf{M}]\times\mathbf{1}_r}{r^2}d\tau. \quad (3\text{–}9)$$

Equation (3–7) represents the field of a continuous volume distribution of magnetic dipoles. It is based on Coulomb's law [Eq. (3–5)] and the dipole concept. Equations (3–8) and (3–9) are derived from Eq. (3–7) by direct transformation.[48] The former equation represents the same field as Eq. (3–7) computed from the net volume density of the magnetic pole strength, $-\mathrm{div}\mathbf{M}$, while the latter represents the field as the sum of contributions from the uncancelled Amperian current density $(1/\mu_0)\mathrm{curl}\mathbf{M}$. For the purpose of computing **H** or **B** at *exterior* points all three formulas are equally legitimate.

It should be emphasized at this point that, although the field of a distribution of polarization is the same at exterior points whether we compute it from the equivalent magnetic pole distribution, or from the equivalent distribution of uncancelled Amperian currents, this is by no means the case for interior points. In the case of a permanent bar magnet *the fields computed in these two ways have opposite directions at interior points. Hence arises the need for distinguishing between two different magnetic vectors at interior points of a magnetic medium.*

The magnetic intensity **H** can now be defined as a theoretical concept for interior points of a magnetic medium if we specify that it shall be a vector quantity composed of two terms, $\mathbf{H}'$ and $\mathbf{H}_s$, to be compued as follows: $\mathbf{H}'$ is the external intensity due to currents or distributions of magnetic polarization outside the medium and computed as if the medium were not magnetic; $\mathbf{H}_s$ is the self-field of the medium computed according to Coulomb's law for a vacuum from the distribution of pole strength which accompanies the magnetization of the medium. In other words $\mathbf{H}_s$ is to be computed from Eq. (3–8). Thus, in an electromagnet **H** is the resultant of the field of the exciting coil in air and the field of the induced distribution of pole strength at the ends of the core, if any. This definition makes **H** a magnetic analog of **E** (see Secs. 2.4 and 2.5).

[48] Equations (3–7), (3–8), (3–9) are equivalent to Eqs. (11), (13), and (12) in "Field vectors and unit system in the theory of electricity," William Fuller Brown, Jr., *Am. J. Physics* **8**, 338 (1940).

The magnetic induction **B** can be defined in turn for interior points of a magnetic medium by the equation

$$\left.\begin{array}{ll}\text{(Gauss)} & \mathbf{H}'+\mathbf{H}_{sa} \\ \text{(Giorgi)} & \mu_0(\mathbf{H}'+\mathbf{H}_{sa})\end{array}\right\}=\mathbf{B}, \quad \begin{array}{r}(3\text{–}10a)\\(3\text{–}10b)\end{array}$$

where $\mathbf{H}_{sa}$ is the magnetic intensity of the medium computed from its uncancelled Amperian currents, that is, on the basis of Eq. (3–9).

State, or prove, that the definitions of **H** and **B** at interior points lead to Eq. (2–43) and call attention to the parallelism between this equation and Eq. (2–42) for electrostatics. Note that Eq. (2–43) can be used as an alternative definition of either **B** or **H** in terms of the other. On the basis of Eq. (2–43) and the theorem of Gauss show that the flux of the outward component of **B** from any closed surface is zero, there being no magnetic equivalent of conduction charge. (In advanced courses it will suffice to show that the divergence of the right hand member of Eq. (2–43) vanishes identically.)

Call attention to the experiments of Barnett and others which prove that angular momentum must be supplied to the elementary magnets when iron is magnetized, thus confirming the Amperian current theory.

Explain that if the Amperian current theory is correct one would expect that the induced electromotive force in an electromagnet would be determined by the rate of change of the flux of **B**, rather than by the rate of change of the flux of **H**. Show further that in fact the generalization of Faraday's law of induction [Eq. (3–6)] for circuits near magnetic bodies requires that the flux Φ be calculated from a vector whose lines are all closed loops (zero divergence), such as **B**. Explain that these considerations are supported by experiment and that the theory identifies the flux of our conceptual vector **B** with the experimental flux through an electromagnet derived by the generalization of the law of electromagnetic induction to include magnetic cores.

The magnetic intensity **H** is important in the ideal case of a ring-shaped electromagnet with a continuous soft iron core because **H** is identical in this case with the external field which causes the magnetic polarization. It is convenient to regard **H** as the "magnetizing force" even when

we are considering the magnetization of a short iron bar, though this interpretation must be regarded as a convention.

The Kelvin-cavity procedures for the "experimental" measurement of **B** and **H** can be mentioned at this point, but the difficulty of carrying out these procedures should not be glossed over. Note that there is no general way to determine experimentally the magnetic polarization of a hard magnet of unknown previous history, although the average changes in **B** can be observed by induction. Call attention to the fact that **B** is the average of the microscopic field—a fact which harmonizes with the Kelvin method of measurement, for the removal of a transverse slot does not affect the distribution of uncancelled Amperian currents which produces the field to any appreciable extent, whereas the cutting of a longitudinal drill-hole required for measurement of **H** produces additional uncancelled Amperian currents just sufficient to account for the difference between **B** and **H**. Observe that on the other hand **E**, not **D**, is the mean microscopic electric intensity and that in the electric case, where the fields are due to charge densities, it is the transverse slot which produces new polarization charges and accounts for the fact that **D** differs from the average microscopic field. This can naturally lead to a discussion of the antiparallel correlation of **E** with **B** and of **D** with **H** which we have considered in Sec. 2.6. It should be made clear that the correlation of **H** with **E** is a convenient device of elementary instruction and of little physical significance, whereas the correlation of **B** with **E** is of fundamental importance.

At this point the instructor can turn to the discussion of the experimental relation between **M** and **H**. Describe the behavior of paramagnetic and diamagnetic materials and introduce the hypothesis that in such materials the magnetization is everywhere proportional to the magnetic intensity **H**. Introduce the magnetic susceptibility χ_m and the relative magnetic permeability κ_m as parallels of the electrical susceptibility and the relative dielectric constant, respectively. (Cf. Sec. 2.4.) Show that in a magnetic fluid, where the measurements can be made, the energy principle and the law of electromagnetic induction require that the current-magnetic-field force shall be BIL for a transverse field and *not* HIL (Gauss) or $\mu_0 HIL$ (Giorgi). Explain that the effect of the permeability of the medium on the current-magnetic-field force is due to uncancelled Amperian currents which appear in the medium at the surface of any current-carrying wire which passes through it and to forces originating in the magnetostriction of the fluid.

Discuss the atomic explanation of the difference in behavior of paramagnetic and diamagnetic materials. Then proceed to the more complex phenomena of ferromagnetism. Give an account of the magnetic properties of ferromagnetic materials in terms of the magnetization curve and the hysteresis loop. Explain the existence of permanent magnets in terms of hysteresis. (See Appendix B.) Define permeability for the ferromagnetic case and point out that the chief use of the quantity is to simplify calculations in which the magnetization curve, or a part of it, is regarded as a straight line, $\mathbf{B}=\mu\mathbf{H}=\mu_0\kappa_m\mathbf{H}$, or $\mathbf{B}=\mathbf{B}_0+\mu\mathbf{H}$, as a rough approximation. Apply the theory to the magnetic circuit.

3.3 Outline II. *The Amperian Approach*

3.31 The Electrostatic Field in Free Space.—On this subject there is no essential difference between the present approach and the traditional approach described in Sec. 3.21. (Section 3.22 on dielectrics, also will be incorporated into this approach, but after the next few sections, which deal with forces between current-carrying wires in air.)

3.32 The Measurement of Currents.—Call attention to the fact that all forces discussed thus far have been between static charges of electricity, so that the laws stated have neither affirmed nor denied the possibility of other forces existing when electricity is flowing. As the simplest apparatus for showing such new forces, consider a pair of straight, parallel wires, of length much greater than the distance between them. Let one of them contain a readily movable segment (flexible, or with flexible links, or with mercury cups at its ends) on which forces may be demonstrated and perhaps measured.

Before using any current, demonstrate the forces due to static charges on the wires, recalling the general law of mutual repulsion between like charges and attraction between unlike.

Now introduce the idea, familiar to most students before the course began, that an electric current is a flow of something that is conserved. Use this idea to show that a simple U-turn connection will ensure opposite directions for currents in these wires, whereas certain other connections will give "like" currents. Then demonstrate that "like" currents cause attraction and "unlike" currents cause repulsion.

Contrast this law with the law for charges, and forecast similar antiparallelism in other laws to be studied soon.

In preparation for expressing this contrast algebraically, call attention to the need for a quantitative definition and measure of current. List a few effects which might be used as measures, linear or nonlinear. The list may include heat, electrolysis, and perhaps other effects, in addition to the forces just observed. In any case, it should include the ability of a current to carry charge to or from a body. This property will have been observed by the class, both in sparks and in the conduction of charges from influence machines through wires to other bodies. Moreover, the atomicity of charge, shown by Millikan's oil-drop experiment (Sec. 3.21), gives special importance to the measurement of current by the time-rate of transport of charge.

Remind the class of the methods used in measuring the flow of water, gas, and heat. Define the strength of an electric current in harmony with the definitions of other currents as quantity of charge-per-unit-time transported across a section of the conductor in which the flow occurs. Note that in the case of steady flow in an unbranched circuit the current will be the same for all sections. Refer to the equivalence of the flow of positive charge in one direction and of negative charge in the opposite direction (see Sec. 3.24).

Promise that any practical device adopted for measuring current will eventually be proved to conform to the foregoing definition. And remember to keep this promise at an appropriate time, by checking the proportionality of some alternating currents, as measured electromagnetically, to the potential differences created by them in a condenser. Indeed, it may be well to suggest now, without mathematical details, how the promise will be kept.[49]

Note, also, the lack of any simple way to measure electrostatically a direct current as strong as the weakest current used in showing the forces between the parallel wires. Note that ordinary lamps and motors require currents of this order, and illustrate this requirement by introducing a bank of parallel lamps in series with the wires. Then propose the problem of deriving a measure of current from the forces on such wires.

In discussing this problem, recall the mutual independence of electrostatic forces between elementary particles, indicated by the product Q_1Q_2 in electrostatic force equations. Postulate, therefore, a similar independence of electromagnetic forces between elementary particles, and a corresponding product I_1I_2. Note that the law of conservation of charge requires direct currents in series to be equal; and draw the conclusion—subject to the above promise of verification—that with a series connection the force on the movable wire is proportional to the square of the current.

Now define *the ampere* as the current which can cause just 2×10^{-7} newton per meter of length in either of two infinitely long wires, connected in series but geometrically parallel, just one meter apart.

Note that the number 2×10^{-7} resulted historically from a round-about definition now superseded. Note also that Ampere himself discovered laws (see Sec. 3.34) through which we can calculate relations between forces on wires in different configurations; so we can make standardizing apparatus much smaller and handier, and more accurate in practice, than these wires many meters long.

As a first example under Ampere's laws, recall the two allusions to the "meter" in the definition of the ampere, and note that if we had said "cm" in each case, or even "inch," we should still have said "2×10^{-7} newton." This fact brings the design of a current balance, accurate

[49] Ideally, of course, this suggests an absolute current balance and an absolute electrostatic voltmeter. Practically, all but the most stubborn skeptics in any class will be convinced and enlightened by a demonstration with an ordinary ammeter and electrostatic voltmeter, known to be reasonably well calibrated. Even an ordinary voltmeter will do, if supported logically. So there is a wide range of choice.

Incidentally, the constant of proportionality may be measured, at least roughly, by using a condenser whose capacitance is geometrically calculable. Naturally, such a capacitance is small. Even with a milliammeter, this measurement requires either a high frequency or a high voltage. Which to use may be a question of taste or availability. With a small x-ray transformer the measurement is easy at 60 cycles. A square meter of glass, 1 cm thick, with tinfoil, will serve as the condenser. An ordinary voltmeter is a good milliammeter. A spark gap between spheres is a voltmeter whose electrostatic nature is spectacularly obvious—too obviously spectacular unless precautions are taken to protect the milliammeter.

enough to be convincing, within reach of an elementary class.[50]

This fact also gives the algebraic form of the force law for such wires; so it enables us now to *formulate the contrast between their electrostatic and electromagnetic forces in the antiparallel equations given in Sec. 2.6 as Eqs. (2–33) and (2–34).*

Describe next the Kelvin current balance, stating that its relation to the straight wires has been calculated through Ampere's laws—not asking freshmen to follow the calculation. How far to go into detail on this and more modern current balances is a question of taste; the main point which the student should learn is that with sufficient mathematical and technical skill the ampere can be reproduced accurately enough to give really concrete meanings to equations about currents.

Next, note the possibility of experimentally calibrating a current balance for which the calculation is too difficult, by running it in series with a calculable one. Then describe the electrodynamometer type of balance, most often used in wattmeters and a.c. voltmeters.

Finally, on ammeters, note the existence of other types. Give some description of them, according to taste and available time; and emphasize the dependence of most of the common, portable ammeters on experimental calibrations, leading back ultimately to some calculable current balance and the definition of the ampere.

3.33 D.C. Circuits.—At this point we are free to choose between at least three programs, according to taste or circumstances, such as the needs of the laboratory. We may go on with the discussion of forces; we may consider electrostatic and magnetic fields in material media; or we may take up the theory of d.c. circuits.

For d.c. circuits, either now or later, this approach includes most of Sec. 3.24 of the traditional approach, entitled "Electric Currents." Some topics in that section, to be sure, have been covered already in this approach; and some others should be postponed until after discussions of fields in material media; but many can come now.

Among these are the following:

(1) Heating effect and its relation to current;

(2) The importance of potential differences, overshadowing though not eliminating that of field strength;

(3) The fact that when a conductor carries a current it is no longer equipotential, and the electrostatic field is no longer normal to its surface;

(4) Electrolysis and Faraday's law;

(5) Chemical electromotive forces, and the reasons for distinguishing them clearly from potential differences;

(6) Ohm's law for conductors not containing emf's, with a warning that the law does not apply to gases and that there are emf's other than chemical;

(7) The relation of the terminal potential difference of a battery to its emf and current, and Ohm's law for a whole circuit;

(8) The custom of connecting household appliances in parallel, and the reason for it;

(9) Resistances in parallel and in series;

(10) Perhaps even Kirchhoff's laws, though it may be better to postpone them.

3.34 The Magnetic Fields of Steady Currents.—Recalling the force laws for straight, parallel wires, raise the question of forces between other circuits. Have a large circular coil on the lecture table, mounted with its plane vertical. (A coil used habitually by the writer of this outline has a diameter of 1 foot and a cross section of 1 square inch, with enough turns to permit the use of 120 volts d.c. for a few minutes at a time.) Have a piece of flexible wire on a handy frame, like a violin bow with loose strands, except for appropriate insulation; and have connections for sending plenty of current through the flexible

[50] For proof of this statement we may cite the experience of R. D. Richtmyer and W. W. Hansen, who described an excellent current balance of this sort in the *American Physics Teacher* **7**, 52 (1939). In this balance the movable wire is hung from a chemical balance. It is horizontal and is surrounded by four parallel fixed wires, so spaced and connected as to make the field as uniform as possible near the movable wire. A lecture-room current balance may be made with only three wires, all on a horizontal plane. In one well-tested balance of this sort, the outer wires are 3 cm apart, and the middle wire has a movable section 10 cm long, hung from 15-cm threads, with its ends dipping into mercury cups 1.5 cm in diameter. Essential parts are projected optically, and the magnification is computed from the distance between the images of the outer wires. Unless the mercury is very clean its surface scum (even if invisible) necessitates tapping the base board, to free the movable wire. This may limit the accuracy to a few percent, but the simplicity of this apparatus makes it demonstrate the principle very clearly.

wire. Hold this bow so that the flexible wire is in the plane of the coil, in various places and orientations, and show how the law about parallel wires offers suggestions about the directions of the forces which will move the flexible wire when both currents are turned on. Then verify these suggestions.

(The writer uses d.c. with separate switches; and he ballasts the flexible wire with a bank of tungsten lamps, to make it jump vigorously when its current is turned on with the lamps cool. But if d.c. is not available, a.c. will do, provided the flexible wire is connected in series with the coil, to keep their currents in phase.)

Next, hold the bow so that the midpoint of the flexible wire is at the center of the coil, and alter its direction to show the changes in the magnitude of the force as well as the force-reversal which accompanies a reversal of the current. Show that there is only one line through this point, along which the flexible wire can lie without experiencing any force. Define the direction of the magnetic field here as that of this line, looking along it in the sense which makes the current in the coil clockwise.

To explore further, replace the flexible wire with a small coil, painted yellow on one face and blue on the other, with current fed to it through a good, flexible lampcord. Carrying it by this lampcord, hang it in the center of the large coil, use the law about parallel currents to predict its ability to identify the direction of the field; then verify this prediction.

Now move the small coil while the currents are on, showing the continuity of its identification of direction. Under its guidance, make a rough map of the lines of the field. Show that these lines are closed loops. Contrast this characteristic with the plus-to-minus, never-looped characteristic of electrostatic lines.

Make the map more complete and accurate by laying a flat, white, horizontal platform through the center of the coil and sprinkling it with iron filings. In this plan of teaching, this use of iron filings is frankly empirical.

With filings and the small coil, explore the fields of some other circuits: perhaps a square coil; a straight solenoid, in the form of a helix of thick wire with turns far enough apart to show the map within, or even to let the coil slip between them; perhaps also, a ring solenoid of similar construction; and so on. In all cases, note the closed-loop characteristic of the lines, contrast this with the plus-to-minus characteristic of electrostatic lines, and record this contrast as a second example of the antiparallelism of the systems of laws.

Measurement of Magnetic Fields.—The differences in strength of field from place to place appeared qualitatively with the exploring coil. To make these differences quantitative *define the magnetic induction* **B** in free space either in terms of the torque on a standard exploring coil, held perpendicular to the plane it likes best, or better, through the equation

$$\text{(Giorgi)} \qquad F = LIB\sin(\mathbf{I},\mathbf{B}), \qquad (3\text{–}11)$$

or possibly

$$\mathbf{F} = L\mathbf{I}\times\mathbf{B}.$$

If using Gaussian units, introduce the name as an abbreviation for the combination "dyne/cm abamp." If using mksc, on the other hand, there is no gain in changing at this point from "newton/meter ampere" to "weber/square meter."

Field Strengths near Circuits.—From this definition of **B** and the force law for parallel wires, deduce the equation for **B** near a long, straight wire.

Note the inverse proportionality of this **B**, at different points, to the circumferences of the lines of **B** through these points; then extend this inverse proportionality law, modified to apply to line-average values of **B**, to lines linked with other circuits. Demonstrate it qualitatively by differences in the liveliness of the small coil.

Demonstrate, also, that wherever lines of **B** diverge, **B** gets weaker. Recall the similar law for **E**. Then state the law for **B** quantitatively like that for **E** in free space; but note also the contrast, that for **B** there is never any net flux through any closed surface.

Apply this law also to the ring solenoid, deriving the equation for its B. Then note that the infinitely long, straight solenoid can be regarded as the limit approached by one side of a ring solenoid which has become infinite in the proper way. In this way find the equation for B in the straight solenoid.

Note that the force on a wire lying inside the solenoid could not be calculated directly from

the law about long, straight, parallel wires; that indeed the extensions of field laws just outlined have really been extensions, even though each step was suggested by something previously known.

To complete the theory, therefore, state the law (for currents in free space)

$$\text{(Gauss)} \qquad \mathbf{B}=\frac{I}{c}\oint\frac{d\mathbf{s}\times\mathbf{1}_r}{r^2}, \qquad \text{(3–12a)}$$

$$\text{(Giorgi)} \qquad \mathbf{B}=\frac{\mu_0 I}{4\pi}\oint\frac{d\mathbf{s}\times\mathbf{1}_r}{r^2}, \qquad \text{(3–12b)}$$

noting, of course, that it is one of several mathematically equivalent forms of Ampere's laws (see Sec. 3.32). Illustrate it with the field on the axis of the circular coil. If using calculus, check it by calculating the fields of the straight wire and the solenoid; or in any case, show that it looks reasonable in relation to these known fields.

Finally, note and demonstrate briefly that the introduction of any large piece of iron causes great changes in a magnetic field; but postpone further discussion of these effects until after that of the corresponding effects in electrostatics.

3.35 Polarizable Media.—(a). *The Electrostatic Field in Material Media.* On this subject, see Sec. 3.22.

(b). *The Effects of Iron on Magnetic Fields.*—Section 3.22 ends with the students' minds on molecules, so it may be well to keep them there in starting the discussion of iron. Static charges confined within molecules enabled us to explain all **E**'s on one basis, without recourse to any modification of the law of force, like the old "weakening of the ether." So currents, similarly confined, enable us to unify all **B**'s.

Bring in some of the history of the concept of Amperian currents. Tell how Barnett and others found angular momentum associated with magnetism. Note the unexpected value of the ratio of angular momentum to magnetic moment and indicate its interpretation in terms of electron spin. But warn students that later developments preclude any literal acceptance of the picture of a spinning sphere of charge.

Summarize the theory by the statement that pieces of iron and other ferromagnetic and paramagnetic materials affect fields in adjacent parts of free space *as if* their atoms contained circuits behaving like a coil carrying a current.

To test the theory on the lecture table, try first a permanent bar magnet. Note that the alignment of its Amperian circuits should make it equivalent to a solenoid; and with iron filings, show the similarity of their external fields. Note also that in the field of the big coil the bar should orient itself just like the exploring coil; then show that it does.

For the theory of electromagnets, recall the fundamental place in the theory of dielectrics, taken by the infinite parallel-plate condenser, and introduce as its magnetic counterpart the infinite straight solenoid. Recall how and why the polarization charge on each surface of a dielectric plate was *opposite* in sign to the adjacent conduction charge. In contrast, show that if individual Amperian circuits behave like the exploring coil, the introduction of an iron core into the solenoid will result in the formation of a net polarization current in the surface of the core in the *same* direction as the adjacent conduction current.[51]

Show similar magnetic effects for a variety of cases. Note that whenever the space within a coil is fairly well filled with iron, the iron is practically equivalent to an Amperian current running alongside the conduction current. Well magnetized iron is equivalent to about a million ampere-turns per meter.

This rule greatly facilitates qualitative reasoning about field, by enabling us to deal with a single field-producing agent instead of two.

3.36 Mechanical Forces with Iron.—Illustrating further the convenience of this way of looking at iron, show the parts of a few motors (of almost any modern kind). Note how thoroughly the

[51] For a test, have a suitable condenser and solenoid on the lecture table. The condenser plates are best mounted in vertical planes, a few centimeters apart, on good insulators. Connect them to an electrostatic voltmeter (say a 3000-volt Braun, or an electroscope on the projection lantern) and charge them so as to get a high reading. Make sure the insulation is good enough to keep the voltmeter reading from decreasing too fast after the source of charges is disconnected. Now, without touching the condenser plates, insert a heavy glass plate between them, noting the quick drop in potential difference; then withdraw the glass, noting the rise. Then make a corresponding test with the solenoid. Hang the little exploring coil reasonably near one end of it, to indicate changes in its external field strength. While the currents are on, insert the iron core and withdraw it. Contrast the rise in magnetic field strength caused by the iron with the drop in electric field strength caused by the glass. (The electrostatic part of this test is Experiment E-70 in Sutton's *Demonstration Experiments in Physics* (McGraw-Hill, 1938).

iron fills most of the space enclosed by the conduction circuits. Note especially that the moving conductors which come nearest to stationary conductors are embedded in slots in the iron; and in many motors these stationary conductors are likewise embedded.

Moreover the conductors of these two groups are usually parallel. This brings the main part of the driving force qualitatively under the simple rule for parallel wires, with each wire vastly reinforced by the Amperian circuits alongside it.

Lifting magnets also show the use of such reinforcement of currents; and the things lifted by them illustrate the alignment of Amperian circuits, still like the exploring coil, but not so well surrounded by conduction currents or other Amperian currents already aligned.

Quantitative calculations on such forces are rare in elementary courses, but the following problem deserves attention, at least in a second course. This is the problem of flat surfaces of iron in contact. Calculus is needed here, whether we think in terms of currents or poles, unless we use the Faraday-Maxwell stresses. We need even more calculus to prove that the stresses give correct results, in general; but this may be a good case with which to introduce them if we want to.

Without them, the comparison of pole surfaces to condenser plates is more obvious to us teachers, who were brought up on poles, than its equivalent in Amperian currents. However, a pair of similar longitudinally magnetized rods with flat ends is equivalent to a pair of solenoids. The force acting on solenoid 1 depends on its current and the field due to solenoid 2. All the lines of this field B_2 that enter the end of solenoid 1 must leave it through its sides, if it is long enough for an analysis in terms of poles to neglect its other pole. And if $n_2I_2=n_1I_1$, half the flux through the surface of contact is that of B_2. So the calculation is easy after all.

3.37 Fields in Iron.—The fundamental problem of defining **B** and **H** for interior points of ferromagnetic materials was considered in Sec. 3.25, but from a different approach. Introduce the subject by defining **B** theoretically for interior points of a magnetic medium as a vector quantity computed from the total distribution of conduction currents and uncancelled Amperian currents exactly as **B** is computed in free space, i.e., in accordance with Eq. (3–10).

Following the procedure used in electrostatics, turn next to the magnetic polarization **M**, defined as the resultant vector moment-per-unit-volume due to the Amperian currents. For simplicity it may be well to base the discussion on the idealized case of a cylindrical bar uniformly polarized along the axis, being sure to indicate later the need for generalization of the conclusions. In this special case the uncancelled Amperian currents are obviously confined to the lateral surface and the current flowing around a slice of thickness dL and area S perpendicular to the axis is given by the formula for the magnetic moment of a plane current loop, in Giorgi-Sommerfeld units $\mathfrak{M}=IS$. Hence the surface density of the current dI/dL is Giorgi-Sommerfeld)

$$\frac{dI}{dL}=\frac{d\mathfrak{M}}{SdL}=M. \qquad (3\text{–}13c)$$

Introduce **H** by Eq. (2–44), which in the Sommerfeld units takes the form $\mathbf{H}=\eta_0\mathbf{B}-\mathbf{M}$. This equation shows that outside the magnet **H** is identical with **B** except for the constant factor η_0, whereas within an electromagnet the term $-\mathbf{M}$ makes **H** much weaker than $\eta_0\mathbf{B}$, and within a permanent magnet not aligned by an external field this term usually gives **H** a direction more or less opposite to that of **B**. In the case of a permanent magnet that is not subject to external fields due to currents, the lines of the vector **H** lead from one end of the bar to the other, both inside and outside the magnet. This characteristic is like that of the lines of **E**, that lead from positive charges to negative, and suggests the possibility of computing **H** from distributions of positive and negative pole-strength at the two ends, just as we could compute **E** from distributions of positive and negative charge. Identify the product of the pole-strength of a permanent magnet and the distance between the poles with its total magnetic moment. Then note the existence of a rigorous method of computing **H** when **M** is given from poles created by the termination of the lines of **M**. Finally, call attention to the fact that in the case of an infinite solenoid with an iron core, or a ring solenoid with a core, the absence of poles means that **H** reduces to the contribution of the conduction currents alone, independent of the magnetization.

Consider next the experimental methods of

determining **B** and **H**, which must, of course, agree with the calculated **B** and **H** wherever comparison is possible. In free space an exploring coil measures either **B** or **H**; but in magnetizable fluids it measures **B**. This is true whether we calculate the field from the mechanical torque on a coil at rest, as described heretofore, or from an electromotive impulse when it is flipped. The electromotive force law is, of course, available to give us average values of **B** even in the interior of iron or steel specimens. The discussion of this law can follow the lines laid down in Sec. 3.24 and might well be taken up before the discussion of magnetic materials.

It is of interest that the **B** measured by the torque or flip-coil methods in a magnetic fluid is the external **B′** of Eqs. (2–21) to (2–23), even if the coil is made of iron wire. The fact that the electromotive impulse measures this **B** is unquestionable. And the principle of the conservation of energy demands that both forces measure the same **B**. So it is not surprising that a reasonable use of the concept of Amperian currents gives a clear explanation of the mechanical force of Eq. (3–11) even when the surrounding medium, or the wire, or both, are magnetizable. Nor is it surprising that in *this* explanation the force density within the wire may be considered to be **J**×**B**, where **J** is the current density and **B** is the **B** inside the wire.[52]

It does not seem profitable to go into details on these forces for elementary students, because after all, magnetizable fluids are too weakly magnetizable to hold the attention of an elementary class much longer than it takes to play with liquid oxygen and a pointed electromagnet—even for this time the students are not attracted by the magnet as much as by that spectacular liquid. Advanced students, however, seem well interested in the analysis of these forces and in the related problem of the unipolar motor, despite the fact that the latter looks much farther from engineering practice than it did thirty years ago. When electrons flow through a bar magnet, from a slip-ring around its waist to a contact point in the middle of one end, the torque is calculable from a force-density **J**×**B**, regardless of whether **H** is even in the same direction as **B**. Thereby hangs a tale—well told by *Zeleny and Page*.[53]

On magnetization curves and the phenomena represented by them there is little to be said here. To be sure, we have to explain as historical the custom of using **H**, rather than **B**, as the independent variable; but its relation to the infinite solenoid makes it empirically independent there, and in the ring solenoid on which most magnetization curves are based.

Mention paramagnetism and diamagnetism giving due emphasis to the smallness of their susceptibilities and to their lack of practical importance. Explain their basic theory qualitatively, and note that diamagnetism would be very hard to account for except as an effect of induced emf's on what are essentially Amperian circuits.

Appendix A: Derivation of Formulas for the Total Force and Torque of Electrical Origin Acting on a Charged Body Immersed in a Fluid Dielectric and Subject to an External Electric Field (Cf. Note 13)

Consider the general case of a body X immersed in a Class A dielectric and carrying arbitrary distributions of polarization charge and conduction charge. It is not necessary that X be homogeneous or linear (in the sense that the polarization is proportional to the electric intensity). Part of the polarization may be permanent. To prove the validity of Eq. (2–11) we set up a modified problem involving a body X^* similar in size and shape to X and subjected to the same external electric field. Body X^* is to have a uniform dielectric constant equal to $(\kappa_e)_A$ and a distribution of conduction charge ρ_c^* so contrived that the total charge density $\rho_t^*=(\rho_c^*+\rho_p^*)$ is equal to the total charge density ρ_t in the original problem of the body X. Then $\rho_c^*=(\kappa_e)_A\rho_t^*=(\kappa_e)_A\rho_t$. Since the electrical intensity **E** depends solely on the external and internal distributions of total charge density, it will be the same in the modified problem as in the original one. To compute the total force $\mathbf{F}_e^*$ for the modified problem we make use of Eq. (2–10) in which the term in gradκ_e now drops out. Then

$$\mathbf{F}_e^*=\int_{X^*}\mathbf{E}\rho_c^*d\tau=\int_{X^*}\mathbf{E}\rho_t^*(\kappa_e)_A d\tau$$

$$=(\kappa_e)_A\int_X \mathbf{E}\rho_t d\tau. \quad \text{(A–a)}$$

(This result can be established independently of Eq. (2–10) by reference to the fundamental theorem regarding the energy of a distribution

[52] See J. A. Stratton, *l.c.*, p. 158, Eq. 13.

[53] J. Zeleny and L. Page, *Physical Rev.* **24**, 544 (1924). This does not mean that these phenomena prove Ampere's theory.

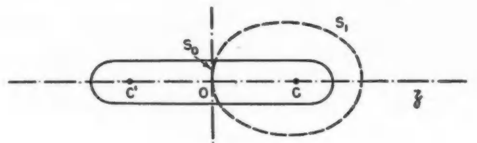

FIG. B1. Theorem of Gauss applied to bar magnet.

of conduction charge in an infinite Class A dielectric.) Because the total charge density and the electric intensity **E** are the same in X and X^*, and because the conditions throughout the fluid are unchanged as we pass from one problem to the other, it is evident that both the purely electric and the electrostrictive contributions to the total force must be the same for the modified problem as for the original problem. Hence $\mathbf{F}_e^*$ is the same as $\mathbf{F}_e$ and Eq. (2–11) follows directly from Eq. (A–a).

It is easy to carry through a similar argument showing that the torque acting on X due to purely electric and electrostrictive forces is the same as the corresponding torque for X^*. Hence we obtain a torque equation analogous to Eq. (2–11) and applicable to any body X immersed in a Class A fluid dielectric, *viz.*,

$$\mathbf{L}_e=(\kappa_e)_A\int_X \mathbf{r}\times\mathbf{E}\rho_t d\tau. \qquad \text{(A–b)}$$

We can proceed in reverse direction to derive Eq. (2–10) from Eq. (2–11) by the following direct transformation, if we introduce the restriction that the material in X is everywhere linear and isotropic, though inhomogeneous, so that the relation $\mathbf{D}=\kappa_e\mathbf{E}$ holds with κ_e a scalar function of the coordinates independent of **E**. In this derivation the indicated surface integrals are to be extended over a surface S immediately adjacent to the surface of X, but inside the surrounding fluid. Then, using Gaussian units

$$\mathbf{F}_e=\frac{(\kappa_e)_A}{4\pi}\int_X(\nabla\cdot\mathbf{E})\mathbf{E}d\tau=$$

$$\frac{(\kappa_e)_A}{4\pi}\left\{\int_S(\mathbf{n}\cdot\mathbf{E})\mathbf{E}dS-\int_X(\mathbf{E}\cdot\nabla)\mathbf{E}d\tau\right\}. \qquad \text{(A–c)}$$

But

$$\int(\mathbf{E}\cdot\nabla)\mathbf{E}d\tau=\frac{1}{2}\int\nabla(E^2)d\tau=\frac{1}{2}\int\mathbf{n}E^2dS.$$

Hence

$$\mathbf{F}_e=\frac{1}{4\pi}\int_S[(\mathbf{n}\cdot\mathbf{D})\mathbf{E}-\tfrac{1}{2}\mathbf{n}(\mathbf{D}\cdot\mathbf{E})]dS$$

$$=\frac{1}{4\pi}\int_X[(\nabla\cdot\mathbf{D})\mathbf{E}+(\mathbf{D}\cdot\nabla)\mathbf{E}-\tfrac{1}{2}\nabla(\mathbf{D}\cdot\mathbf{E})]d\tau$$

$$=\int_X\rho_c\mathbf{E}d\tau+\frac{1}{8\pi}\int_X[2(\mathbf{D}\cdot\nabla)\mathbf{E}-\nabla(\mathbf{D}\cdot\mathbf{E})]d\tau. \qquad \text{(A–d)}$$

If $\mathbf{D}=\kappa_e\mathbf{E}$, the integrand of the second integral in Eq. (A–d) reduces to

$$2\kappa_e(\mathbf{E}\cdot\nabla)\mathbf{E}-\nabla(\kappa_eE^2)=-E^2\nabla\kappa_e. \qquad \text{(A–e)}$$

Thus Eq. (A–d) is transformed into Eq. (2–10).

Appendix B: The Behavior of Permanent Magnets in a Magnetic Fluid (Cf. p. 77)

The purpose of this appendix is to investigate the relation between the point-pole approximation and the behavior of real magnets with particular regard to the fields generated when immersed in a Class A magnetic fluid.

Let us consider an isolated magnet which is geometrically and physically symmetric with respect to each of three planes through its center; these we take to be the coordinate planes. We suppose that the magnetic treatment guarantees a corresponding symmetry of the magnetic state, with the magnetic moment along the positive z axis (cf. Fig. B1). Let the magnet be surrounded either by an infinite Class A magnetic fluid of susceptibility χ_m and relative permeability κ_m, or by free space; the formulas for the latter case follow from those of the former by setting $\chi_m=0$, $\kappa_m=1$.

The magnetization **M** of the magnetic material contributes a volume pole-density $-\mathrm{div}\mathbf{M}$ within the magnet and a surface density M_n on its surface. If there is a surrounding fluid, it contributes an additional surface pole-density $-\chi_mH_n^+$, where H_n^+ is evaluated just outside the magnet surface. For $\chi_m>0$ the poles act as if their effectiveness had been reduced by the surrounding fluid; but the ratio in which the effectiveness is reduced is not, in general, given by as simple an expression as the factor $1/\kappa_m$ of Eq. (2–15).

In order to investigate the effect of immersion on the effective or total pole-strength p_t, we may start from the theorem of Gauss. Referring to Fig. B1 we identify S_O with the middle cross section of the magnet and S_1 with a surface outside the magnet which, together with S_O, encloses the right half of the magnet. Then

$$\left.\begin{array}{lr}\text{(Gauss)} & 4\pi p_t \\ \text{(Giorgi-Kennelly)} & \dfrac{1}{\mu_0}p_t \\ \text{(Giorgi-Sommerfeld)} & p_t\end{array}\right\}=\int_{S_O+S_1} H_n dS. \quad \begin{array}{r}\text{(B–1a)}\\ \text{(B–1b)}\\ \text{(B–1c)}\end{array}$$

In the integral over S_1, H_n may be replaced by B_n/κ_m (Gauss) or $B_n/\mu_0\kappa_m$ (Giorgi). With the integral in this form the surface S_1 may be allowed (because of the solenoidal character of **B**) to collapse until it coincides with S_O. The right-hand member of Eq. (B–1) can then be written[54] as $S_O(\bar{B}/\kappa_m-\bar{H})$ (Gauss) or $S_O(\bar{B}/\mu_0\kappa_m-\bar{H})$ (Giorgi) where the bars indicate average values on S_O. Replacement of $\bar{B}$ by its value in terms of $\bar{H}$ and $\bar{M}$ gives

$$\left.\begin{array}{l}\text{(Gauss or Giorgi-Sommerfeld)} \\ \qquad S_O(\bar{M}-\chi_m\bar{H})/\kappa_m \\ \text{(Giorgi-Kennelly)} \\ \qquad S_O(\bar{M}-\mu_0\chi_m\bar{H})/\kappa_m\end{array}\right\}=p_t. \quad \begin{array}{r}\text{(B–2a, c)}\\ \\ \\ \text{(B–2b)}\end{array}$$

In a vacuum Eq. (B–2) reduces to

$$p_0=S_O\bar{M}_0, \qquad \text{(B–3)}$$

so that the pole-strength is determined by the mean magnetization in the middle cross section. In a fluid only a part $S_O\bar{M}$ of the pole strength is contributed by the magnet material; the rest is contributed by the fluid.

If the magnet is supposed replaced by point poles of strengths p_t and $-p_t$ at the positive and negative centroids, C and C', whose position vectors **R** and $-$**R** are defined by

$$2\mathbf{R}p_t=\int \mathbf{r}dp_t=2\int_{z>0} \mathbf{r}dp_t, \qquad \text{(B–4)}$$

the correct moment is, of course, obtained. This procedure gives the correct field at large distances and the correct torque (cf. Eq. (2–22)) exerted on the magnet in a uniform external field. However, these quantities can be computed even more simply by replacing the magnet by a point dipole. At intermediate distances the value of the point pole approximation depends on the shape of the magnet. If the magnet is nearly spherical, the dipole gives a better approximation to the field than the pair of point poles; for the external field of a uniformly magnetized sphere is rigorously that of a dipole at the center, although the distance between the polar centroids is 4/3 of the radius. For an elongated magnet the point pole approximation becomes more useful.

In discussing the special case of an elongated magnet it is convenient to introduce a scalar magnetic potential V defined, except for a constant factor dependent on the units, as $\int dp_t/r$. Then **H** is $-\text{grad}\, V$. The contribution to the integral from the right half of the magnet may be expanded as a series of descending spherical harmonics by expanding $1/r$ in powers of the coordinates with respect to the positive centroid C; the series converges outside a sphere S' enclosing the right half of the magnet. The leading term is p_t/r_1, where r_1 is the distance of the field point from C, and the dipole term in $1/r_1^2$ is zero because of the relation (B–4). A similar expansion holds for the left half; and at points outside both spheres, and far enough away so that the quadrupole terms in the series are negligible, the two point poles give a good approximation to the field of the magnet, though the dipole approximation may be quite unreliable.

The immersion of a "permanent" magnet in a magnetic fluid changes the effective pole strength and also alters the positions of the pole centroids. If the magnet is long, and if the susceptibility of the fluid is not too large, however, the displacement of the centroids may be neglected. The problem of evaluating the change in the external field of a magnet in the point pole approximation due to immersion is therefore reduced to a determination of p_t/p_0 based on Eq. (B–2).

To derive a general expression for the ratio of the two pole-strengths it is convenient to introduce the "ballistic demagnetizing factor."

[54] In this appendix the mean values of the z components of **B**, **H**, and **M** over S_O recur frequently. The subscript z is in such cases omitted as unnecessary. Thus we replace $\bar{B}_z$ by $\bar{B}$.

For a magnet in free space it is customary to define this factor, N_0, by the equation $\bar{H}_0 = -N_0\bar{M}_0$. The factor N_0 is then determined by the geometry of the magnet and the relative pole distribution. It changes only when the distribution of the total pole strength changes. For our magnet in free space it follows from Eq. (B–3) that $S_O\bar{H}_0 = -N_0 p_0$. Since $\bar{H}$ in Eq. (B–2) is proportional to p_i as long as the relative distribution remains unchanged, the generalization of the definition of N_0 for an immersed magnet is evidently $S_O\bar{H} = -Np_i$. Instead of N one can also introduce a dimensionless demagnetizing factor D independent of units and defined by

$$\left.\begin{array}{lll} \text{(Gauss)} & -4\pi Dp_i & \text{(B–5a)} \\ \text{(Giorgi-Kennelly)} & -\dfrac{1}{\mu_0}Dp_i & \text{(B–5b)} \\ \text{(Giorgi-Sommerfeld)} & -Dp_i & \text{(B–5c)} \end{array}\right\} = S_O\bar{H}.$$

The factor D is, in fact, the quantity usually tabulated. The value of D_0 in free space is defined by the same equations with H and p_i replaced by H_0 and p_0. In general, the two D's will be slightly different, since the poles cannot be expected to have exactly the same distribution before and after immersion in the fluid.

If we eliminate $\bar{H}$ from Eqs. (B–2) by (B–5) and express the susceptibility in terms of κ_m, we obtain, independent of our choice of unit systems, the equation

$$p_i = \frac{S_O\bar{M}}{\kappa_m - D(\kappa_m - 1)}. \qquad \text{(B–6)}$$

If we make the special assumption that the magnet is ideally hard, we may set $S_O\bar{M} = S_O\bar{M}_0 = p_0$ to obtain

$$\frac{p_i}{p_0} = \frac{1}{\kappa_m - D(\kappa_m - 1)}. \qquad \text{(B–7)}$$

As a first approximation valid only for large dimension ratio (length/diameter) and good permanent magnet materials[55] we may use the point pole approximation to evaluate $\bar{H}$ in Eq. (B–5). Using Gaussian units and remembering that both poles contribute to H at the midsection, we set $\bar{H}$ equal to $-2p_i/R^2$. It follows that in these units $S_O\bar{H}$ is equal to $-2p_i\omega$, where ω is the solid angle subtended at the centroid C by the midsection. The corresponding value of D (independent of units) is $\omega/2\pi$. This formula gives admittedly inaccurate results when applied to an ellipsoid or to a cylinder of magnetically soft material,[56] but for long permanent magnets of good materials should be sufficiently accurate for our purpose. If the dimension ratio is 50, for example, and the centroid is at 0.9 of the distance from the middle to the end of the magnet, D is 2.5×10^{-4}. In this case Eq. (B–7) indicates that the $1/\kappa_m$ approximation for p_i/p_0 is a good one.

In the case of a homogeneous, isotropic ellipsoidal magnet magnetized uniformly along a principal axis Oz, the demagnetizing field is uniform throughout the interior and opposite to the magnetization. If an ellipsoidal magnet is magnetized in a uniform applied field, the magnetization and the demagnetizing field will be uniform for all values of magnetization. There are only surface poles, of density $M\cos(n,z)$, and if the magnet is surrounded by a Class A fluid, the poles induced on the surface distribute themselves according to the same law. In this case, therefore, D is independent of κ_m. Numerical values of D for different dimension ratios have been tabulated quite thoroughly.[57] They vary between zero (for an infinitely long circular or elliptic cylinder magnetized longitudinally) and unity (for an infinite plane magnetized transversely). Formulas previously given by Page[58] for spheroids and by Wilberforce[59] for the sphere ($D = \frac{1}{3}$) may be derived from Eq. (B–7) by inserting the appropriate explicit formula for D.

Although these formulas already illustrate the failure of the $1/\kappa_m$ ratio for p_i/p_0 (and therefore, in this case, for the ratio of the external **H** after to that before immersion), they are not complete; for they take no account of the actual properties of permanent magnet materials.

[55] Experimental curves on the flux distribution in cylindrical rods by Bozorth and Chapin (*J. Applied Physics* **13**, 320–326 (1942)) show that for soft materials at low fields the approximation under consideration is inadmissible.

[56] For curves showing D for rods as a function of dimension ratio and permeability, see Bozorth and Chapin, *op. cit.*

[57] J. A. Osborn, *Physical Rev.* **67**, 351 (1945); E. C. Stoner, *Phil. Mag.* **36**, 803 (1945).

[58] L. Page, *Physical Rev.* **44**, 112 (1933).

[59] L. R. Wilberforce, *Proc. Physical Soc. London* **45**, 82 (1933).

The production of a stable state of permanent magnetization requires two steps, initial magnetization by a strong positive field and subsequent stabilization by a relatively weak negative field.[60] Upon removal of the strong field, the point in Fig. B2 that represents the magnetic state traverses the descending magnetization curve $SABCD$ as far as B, where $\bar{H}=-N_0\bar{M}_0$; for an ellipsoid this is the intersection of $SABCD$ with a straight line OBQ of slope $-1/N_0$. Application and removal of a field now takes the representative point around a cycle such as BGB if the field is positive but along a path such as BCP if it is negative. The magnet must therefore be stabilized by subjecting it to a negative field larger than any expected in service. If the stabilization process is represented by BCP, subsequent applications of fields no larger than that used in the stabilization cause traversal of thin loops such as $PECP$. The behavior is almost reversible and is represented approximately by a straight line CE. If the magnetization at P is M_0, then for a small change $\Delta\bar{H}$ of magnetizing force (due either to application of a field or to introduction of other magnetic matter into the vicinity) we may define a differential susceptibility χ_m' of the magnet by

(Gauss or Giorgi-Sommerfeld)

$$\bar{M}=\bar{M}_0+\chi_m'\Delta\bar{H}, \qquad \text{(B–8a, c)}$$

(Giorgi-Kennelly)

$$\bar{M}=\bar{M}_0+\mu_0\chi_m'\Delta\bar{H}; \qquad \text{(B–8b)}$$

and although χ_m' is then strictly a variable that depends on $\Delta\bar{H}$ and on the history after stabilization, to a good approximation it may be treated as a constant and identified with the "reversible susceptibility" at C or P. The reversible permeability κ_m' corresponding to susceptibility χ_m' has values, for modern permanent magnet materials, which range from 4 to 12.[61] Thus κ_m' is small by ferromagnetic standards, but it is still large in comparison with the value for any Class A liquid; in the present calculation, therefore, it must not be neglected.

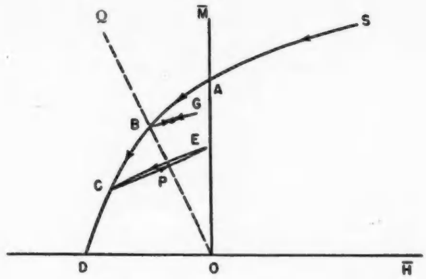

FIG. B2. The magnetization of a permanent magnet.

We may now calculate the ratio p_t/p_0 by the method indicated before, except that instead of equating M to $\bar{M}_0$ we give it the value indicated in Eq. (B–8). The result for all three unit systems is

$$\frac{p_t}{p_0}=\frac{1+D_0(\kappa_m'-1)}{\kappa_m+D(\kappa_m'-\kappa_m)}. \qquad \text{(B–9)}$$

In consequence of our method of defining D_0, D, and κ_m', Eq. (B–9) is rigorous; approximations are introduced only when κ_m' is replaced by a constant and D and D_0 by estimated numerical values. The error involved in the first of these approximations is negligible for our purposes; but D and D_0 can be calculated rigorously only for an ellipsoid. The effect of the magnet permeability κ_m' vanishes when $D=0$ (very long needle-ellipsoid) and when $D=1$ (thin disk ellipsoid magnetized transversely). The ratio p_t/p_0 is $1/\kappa_m$ for the former case and one for the latter; these two results are valid in the limit even if the magnet is not ellipsoidal, since the limiting values of D for the infinite cylinder and for the plane plate hold regardless of the manner of approach to the limiting shape. For other dimension ratios the pole strength ratio p_t/p_0 is intermediate between the values $1/\kappa_m$ for a needle and one for a disk.

In the case of an ellipsoid of dimension ratio 50, D and D_0 have the common value 0.001443. Typically, κ_m' can be set equal to 11 for a hard magnetic material; and for a magnetic fluid κ_m will never differ much from unity. The terms containing D in Eq. (B–9) are small compared with unity and very nearly equal. It is immediately evident that under these conditions the approximation $p_t/p_0=1/\kappa_m$ is a good one.

[60] See T. Spooner, *Properties and testing of magnetic materials* (McGraw-Hill, ed. 1, 1927), Chapts. 2, 5, and 6; especially pp. 11, 59–62, 69–70. See also R. M. Bozorth, *Rev. Mod. Physics* **19**, 29 (1947).

[61] V. E. Legg, *Bell System Tech. J.* **18**, 438 (1939).

For a long magnet of other shape we may use the previous approximation $\omega/2\pi$ as an estimate of D. For the particular case of dimension ratio 50 the numerical estimate is 2.5×10^{-4}. Since κ_m-1 is of the order of 5×10^{-3} for liquid oxygen, we need not consider cases in which κ_m-1 is greater than 0.01. The smallness of D, D_0, and κ_m-1 then permits us to rewrite Eq. (B–9) for the specific class of cases under consideration as

$$\frac{p_t}{p_0}=\frac{1}{\kappa_m}\left[1+D_0(\kappa_m'-1)-D\left(\frac{\kappa_m'-1}{\kappa_m}\right)\right]$$

$$=\frac{1}{\kappa_m}[1+(D_0-D)(\kappa_m'-1)]. \quad \text{(B–10)}$$

To estimate the probable deviation of D from D_0 let us denote the induced pole-strength due to immersion by p_i and the distance from the center of the magnet to the centroid of p_i by R_i. Then

$$p_tR_t=p_0R_0+p_iR_i. \quad \text{(B–11)}$$

Let us postulate that R_i differs from R_0 by 5 percent, a generous estimate for a magnet of the dimension ratio under consideration. Then R_t/R_0 reduces to $1\pm0.05(1-p_0/p_t)$. The ratio D_0/D is equal to $(R_t/R_0)^2$, or $1\pm0.1(1-p_0/p_t)$. Let us now replace $(1-p_0/p_t)$ by the approximation $1-\kappa_m$, thereby reducing D_0/D to $1\pm0.1(\kappa_m-1)$. The quantity $(D_0-D)(\kappa_m'-1)$ in Eq. (B–10) is now seen to be of the order of magnitude of $10^{-4}(\kappa_m-1)$ and thus of negligible importance. We conclude that for long slender magnets of hard steel the approximation $1/\kappa_m$ for p_t/p_0 is very good.

List of Principal Symbols

F_e, $\mathbf{F}_e$ = net mechanical force of electrostatic origin including contribution from electrostriction.
f_e = direct electrostatic force of one charge on another.
q = electric charge (in formulating Coulomb's law).
q_c = conduction charge.
ρ_c, σ_c = volume and surface densities, respectively, of conduction charge.
ρ_p, σ_p = volume and surface densities, respectively, of polarization charge.
$\rho_t=\rho_c+\rho_p$ = volume density of *net* charge.
V = electric potential.
$\mathbf{P}$ = electric polarization.
κ_e = relative dielectric constant.
ϵ_0 = dimensional constant for D/E in free space.
$\epsilon=\epsilon_0\kappa_e$.
χ_e = electric susceptibility (dimensionless).
I = conduction current.
$\mathbf{J}_c$, $\mathbf{J}_m$ = current densities for conduction currents and uncancelled Amperian currents, respectively.
$\mathbf{F}_m$ = net mechanical force of magnetic origin including contribution from magnetostriction.
p = pole strength (in formulating Coulomb's law for poles).
p_0, p_t = pole strength due to hard polarization only and to total polarization, respectively.
ρ_{m0}, ρ_m = pole density due to hard polarization only and to total polarization, respectively.
$\mathbf{M}_0$, $\mathbf{M}$ = magnetic polarization, hard and total, respectively.
μ_0 = dimensional constant for B/H in free space.
κ_m = relative permeability.
$\mu=\mu_0\kappa_m$.
χ_m = magnetic susceptibility (dimensionless).
χ_m' = differential susceptibility of hard magnet.
$\eta=1/\mu$ = reluctivity.
$\mathbf{n}$ = outward normal to surface.

Errata: The Teaching of Electricity and Magnetism at the College Level. I. Logical Standards and Critical Issues

(Report of the Coulomb's Law Committee of the A.A.P.T.)
[*Am. J. Phys.* 18, 1 (1950)]

Equations in the text of Sec. 2.4 second and third lines after numbered Eqs. (2–28a) and (2–28b) should read "$D=\kappa_e\mathbf{E}$" and "$\mathbf{D}=\epsilon_0\kappa_e\mathbf{E}$" respectively.

Equation (2–38a) should be c curl $\mathbf{B}=4\pi(\mathbf{J}_c+\mathbf{J}_m)$.

The Mathematics of Elementary Thermodynamics

Karl Menger
Illinois Institute of Technology, Chicago, Illinois

The purpose of this paper[1] is to bridge the gap between calculus as taught in elementary mathematics courses, and calculus as applied in the first courses on thermodynamics. Many beginners are baffled by two discrepancies.

(A). In his calculus course, the student has, at best, acquired the idea of the "complete" differential df of a function f of several variables. In thermodynamics he learns that the quantity of heat dq, transferred to a gas, is $pdv+du$, where p denotes the pressure, v the volume, and u the internal energy. He is told that dq is an "incomplete" differential. What is an incomplete differential? The answer given in most textbooks of thermodynamics (that it is a differential which is not complete) is incomprehensible to one not familiar with the concept of a general differential, and many presentations of this latter concept are obscure. When clarified, it proves to belong to those ideas which, by virtue of their very simplicity, are beyond a real understanding of beginners (cf. Sec. 8). Some physicists suggest that, in view of the incompleteness of the differential, the quantity of heat should be denoted by δq, $đq$, or Q. This suggestion is good but, obviously, not a substitute for the much needed definitions of the terms used in the basic laws of thermodynamics. In the absence of such definitions, most beginners who give any thought to the equation $dq=pdv+du$ interpret it as follows: If the volume and the energy of a gas undergo small changes Δv and Δu, then the quantity of heat Δq, which has to be transferred to the gas, is approximately equal to $p\Delta v+\Delta u$. As simple examples show, this interpretation is incorrect no matter what we mean by "approximately equal" (cf. Sec. 2).

(B). In a *mathematical* statement, two functions are denoted by the same symbol if, and only if, they are equal: that is, if and only if: (1) they have the same domain, and (2) with each element of this domain they associate the same number or, as we say, the same "function value." For instance, the function associating with each number t, the number $16t^2$, and the function associating with each number s, the number $16+16(s+1)(s-1)$, may be denoted by the same symbol, say F, and we have

$$F(t)=16t^2$$

and $F(s)=16+16(s+1)(s-1)$. In *physics*, two functions are denoted by the same symbol if the meaning of the function values is the same. For instance, the formulas $v=v(t)$, $v=v(s)$ express the fact that the velocity of a falling body is determined by the time t elapsed since the release of the body, as well as by the distance s the body has traversed. Since $v=gt$ and $v=\sqrt{(2gs)}$, the mathematician would use different symbols for the two functions and would write $v=f_1(t)$ and $v=f_2(s)$. In thermodynamics, the student who takes seriously what he has learned in mathematics and in physics runs into difficulties with regard to almost every function. He should see at least one function treated in both notations with a discussion of their relative merits (cf. Secs. 4 and 5).

A corollary of the difference mentioned in (B) is the need for different symbols for partial derivatives in physics and mathematics. The difficulties are enhanced by the use of the same symbol for functions and numbers (variables).

We shall present the first law of thermodynamics without any reference to the concept of a differential (1) as a relation between time rates of change (Sec. 1) and (2) using a Stieltjes integral (Sec. 6). Also the second law will be analyzed without any essential reference to differentials (Sec. 10). An Appendix contains a definition of the thermodynamic equality of quasi-static processes.

1. The Differential Form of the First Law

We begin with a presentation of the first law that presupposes nothing but the idea of the

[1] In preparing the final draft of this paper the author has profited by suggestions from Burton D. Fried and Leo A. Schmidt of the Illinois Institute of Technology, and from Dr. Eric Lype of the Armour Research Foundation.

Part of the content which was presented in lectures at the University of Notre Dame, is incorporated in the Master's thesis of Vincent J. Cushing (1946). An abstract of a paper "The use of differentials in thermodynamics" was published in *Am. Math. Monthly* 56, 210 (1949).

derivative (df/dt) or $f'(t)$ of a function of one variable $f(t)$. We shall study a thermodynamical process Π of a homogeneous substance for which we can determine the following five functions of the time:

$v(t)$, the volume at the moment t;
$p(t)$, the pressure at the moment t;
$\theta(t)$, the temperature at the moment t;
$u(t)$, the internal energy at the moment t;
and
$q(t)$, the quantity of heat transferred to the substance between an initial moment t_0 and the moment t.

We shall call such a process *quasi-static*. An example is the gradual change of volume, temperature, and energy of an ideal gas through the transfer of heat under constant pressure. On the other hand, if by opening a shutter in the container we let the gas suddenly expand into an originally empty neighboring container of equal size, the gas undergoes a process which is not quasi-static. No uniform pressure can be determined while the gas is expanding. The volume is doubled during the process but, at its intermediate stages, the gas is not homogeneous since in the original container the density drops from its initial value δ to $\delta/2$ while in the neighboring container the density rises from 0 to $\delta/2$. Once and for all, we point out that *all processes studied in this paper, except those considered in Statement C of Sec. 10, are quasi-static.*

We call the process Π *differentiable* if the functions $v(t)$, $u(t)$, and $q(t)$ possess continuous derivatives. For a differentiable process, the first law may be introduced as the following relation between $p(t)$ and the time rates of change of the functions $v(t)$, $u(t)$, and $q(t)$

$$(dq/dt)=p(dv/dt)+(du/dt)$$

or

$$q'(t)=p(t)v'(t)+u'(t) \tag{1}$$

at every moment t.

In calculus, the student has learned that if the function $f(t)$ has a continuous derivative, then, for every number t and every number Δt, we have

$$f(t+\Delta t)-f(t)\sim_{\Delta t}f'(t)\Delta t,$$

where the symbol $\sim_{\Delta t}$ indicates that the expressions on the left and on the right side are approximately equal in the following sense. If $|\Delta t|$ is sufficiently small, then the difference between $f(t+\Delta t)-f(t)$ and $f'(t)\Delta t$ is not only small—this is obvious, since both $f(t+\Delta t)-f(t)$ and $f'(t)\Delta t$ themselves are small—but small in comparison with $|\Delta t|$; that is, by choosing $|\Delta t|$ sufficiently small (but $\neq 0$) we can make

$$\left|\frac{f(t+\Delta t)-f(t)-f'(t)\Delta t}{\Delta t}\right|$$

as small as we please.

Applying this theorem to the three derivatives $q'(t)$, $v'(t)$, and $u'(t)$ we see that Eq. (1) is equivalent to

$$q(t+\Delta t)-q(t)\sim_{\Delta t}p(t)[v(t+\Delta t)-v(t)] \\ +[u(t+\Delta t)-u(t)]. \tag{2}$$

Hence the expositor may either start with Eq. (1) and derive Eq. (2) or start with Eq. (2) and derive Eq. (1).

2. A Misinterpretation of the Differential Form of the First Law

In formula (1) which connects $p(t)$ and the time rates of change of the functions $v(t)$, $u(t)$, and $q(t)$, every term is clearly and unambiguously defined. Traditionally, however, the first law is written in the form of an equality of differentials

$$dq=pdv+du. \tag{3}$$

This last equality is clear and unambiguous only when considered as an abbreviation for Eq. (1); in other words, when supplemented by the remark that dq, dv, and du are differentials with regard to the time of the functions of one variable $q(t)$, $v(t)$, and $u(t)$. This point is rarely emphasized. In lieu of the above remark, most books present those discussions of the incompleteness of the differential dq which we mentioned in the introduction to this paper. Experience shows that, in spite of those discussions, many beginners misinterpret formula (3) by translating it into the following statement:

If the volume and the energy of a gas undergo sufficiently small changes Δv and Δu, then the quantity of heat that has to be transferred to the gas is approximately equal to $p\Delta v+\Delta u$. More precisely, Δq and $p\Delta v+\Delta u$ differ by a quantity

which is small in comparison wfth Δv and Δu, or

$$\Delta q \sim_{\Delta v,\,\Delta u} p\Delta v + \Delta u. \qquad (4)$$

While no textbook makes this erroneous claim, few books warn against it. Since, under these circumstances, the misunderstanding frequently occurs, it seems advisable to show explicitly that the above statement (4) is neither a logical consequence of Eqs. (1) and (2) nor generally valid for thermodynamical processes. This can be shown by means of simple examples:

(*a*) *A mathematical example.*—Let $F(x)$, $A(x)$, $B(x)$, and $C(x)$ be four functions satisfying the relation

$$F'(x) = A'(x) + C(x)\cdot B'(x) \qquad (5)$$

for every x. In view of what we saw in Sec. 1, from Eq. (5), it follows that if $|\Delta x|$ is sufficiently small, then

$$\Delta F \sim_{\Delta x} \Delta A + C(x)\Delta B.$$

But the relation (5) does by no means imply that for small changes ΔA and ΔB of $A(x)$ and $B(x)$, the change of $F(x)$ will differ from $\Delta A + C\Delta B$ little, let alone by a quantity which is small in comparison with ΔA and ΔB. If, for instance,

$$F(x) = -(2x-1)^3/6, \quad A(x) = B(x) = x^2 - x, \quad C(x) = -2x,$$

then relation (5) holds and yet $F(x)$ may undergo large changes although $A(x)$ and $B(x)$ change little or not at all. For instance,

for $x=0$ we have

$$F(0) = 1/6, \quad A(0) = 0, \quad \text{and} \quad B(0) = 0;$$

for $x=1$ we have

$$F(1) = -1/6, \quad A(1) = 0, \quad \text{and} \quad B(1) = 0.$$

Hence, as x changes from 0 to 1, the net change of F is $-\frac{1}{3}$, while the net change of both A and B is 0. If x changes from 0 to a number close to 1, then the change of F is almost $-\frac{1}{3}$, while the net changes of A and B are almost 0. It is true that the change of x from 0 to 1 is relatively large and that for smaller changes of x, such as from 0 to $\frac{1}{2}$, the net changes of A and B are appreciable. It is furthermore true that if x changes from 0 to a number which is sufficiently close to 0, then the difference between ΔF and $\Delta A + C\Delta B$ is small in comparison with $[(\Delta A)^2 + (\Delta B)^2]^{\frac{1}{2}}$. But in the erroneous statement (4) one compares the net change Δq merely with the net changes ΔA and ΔB and completely ignores the quantity Δt. This is where the error lies, for the quantity Δt plays an essential role in the correct formula (2).

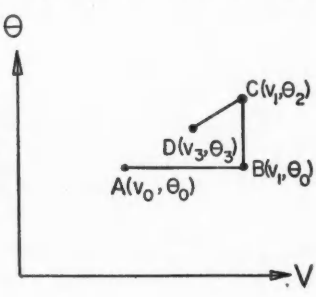

FIG. 1. Thermodynamic process applied to an ideal gas, in which, after appreciable change of q the initial volume and temperature are almost restored.

(*b*) *A physical example.*—We study one mole of an ideal gas for which, at every moment, $pv = R\theta$. The internal energy u remains constant, if the volume is changed while the temperature is kept constant. Morever, u depends linearly upon θ, if either volume or pressure is kept constant. We have

$$u - u_0 = c_v(\theta - \theta_0) \quad \text{and} \quad u - u_0 = c_p(\theta - \theta_0),$$

respectively, where u_0 is the energy at the temperature θ_0 and where the constants c_v and c_p satisfy the equality $c_p - c_v = R$.

Every process in which, after an appreciable change of q, the initial volume and temperature of the gas are restored or almost restored, disproves statement (4). We carry out in detail a simple instructive example due to Mr. Burton D. Fried. One mole of an ideal gas, with the initial volume, temperature, and pressure v_0, θ_0, and p_0, is contained in a cylinder of the type employed in a Carnot cycle, and undergoes the following process Π_1:

(1) First, the gas expands isothermally to the volume $v_1 > v_0$.

(2) Then, the gas is heated at constant volume to the temperature $\theta_2 = v_1\theta_0/v_0$. When it reaches this temperature, the pressure assumes its original value p_0.

(3) Finally, the gas is cooled and its volume decreased under constant pressure. Let v_3 and θ_3 denote the final volume and temperature.

In a (v, θ)-plane, the process is represented by a triangular path from $A=(v_0, \theta_0)$ via $B=(v_1, \theta_0)$ and $C=(v_1, \theta_2)$ to $D=(v_3, \theta_3)$ (see Fig. 1).

Computing Δq in the traditional way we see that $d\theta=0$ on AB; $dv=0$ on BC; and $dp=0$ on CD. Hence,

$$\Delta q=\int_{ABCD} dq=\int_A^B pdv+\int_B^C c_v d\theta + \int_C^D (pdv+c_p d\theta)$$

$$=\int_{v_0}^{v_1}\frac{R\theta_0}{v}dv+\int_{\theta_0}^{\theta_2} c_v d\theta+\int_{v_1}^{v_3} p_0 dv+\int_{\theta_2}^{\theta_3} c_p d\theta$$

$$=R\theta_0 \ln(v_1/v_0)+c_v(\theta_2-\theta_0) +p_0(v_3-v_1)+c_p(\theta_3-\theta_2).$$

Now,

$$c_v(\theta_2-\theta_0)+c_p(\theta_3-\theta_2)=(c_p-R)(\theta_2-\theta_0) +c_p(\theta_3-\theta_2)=c_p(\theta_3-\theta_0)-R(\theta_2-\theta_0)$$

and

$$R(\theta_2-\theta_0)=R\theta_0(v_1-v_0)/v_0=p_0(v_1-v_0).$$

Hence,

$$\Delta q=p_0\Delta v+c_p\Delta\theta+R\theta_0 \ln(v_1/v_0)-2p_0(v_1-v_0),$$

where we have set $\theta_3-\theta_0=\Delta\theta$ and $v_3-v_0=\Delta v$.

If we choose v_3 and θ_3 closer and closer to v_0 and θ_0, then Δv and $\Delta\theta$ approach 0, while the corrective term

$$\Delta q-p_0\Delta v-c_p\Delta\theta=R\theta_0 \ln(v_1/v_0)-2p_0(v_1-v_0)$$

not only fails to become small in comparison with $\Delta\theta$ and Δv but does not become small at all. For the corrective term is independent of v_3 and θ_3 and, as is easily seen, negative provided that $v_1>v_0$.

It may be of interest to describe the process Π_1, by five functions of the time $v(t)$, $\theta(t)$, $p(t)$, $u(t)$, and $q(t)$. Corresponding to the three stages of the process, each of these functions is defined in three time intervals:

$$v(t)=\begin{cases} v_1(t) & \text{for } t_0\le t\le t_1 \\ v_1 & \text{for } t_1\le t\le t_2 \\ v_3(t) & \text{for } t_2\le t\le t_3,\end{cases}$$

where $v_1(t)$ and $v_3(t)$ are functions satisfying the conditions

$$v_1(t_0)=v_0,\quad v_1(t_1)=v_1,$$
$$v_3(t_2)=v_1,\quad v_3(t_3)=v_3;$$

$$\theta(t)=\begin{cases} \theta_0 & \text{for } t_0\le t\le t_1 \\ \theta_1(t) & \text{for } t_1\le t\le t_2 \\ p_0v_3(t)/R & \text{for } t_2\le t\le t_3,\end{cases}$$

where $\theta_1(t)$ is a function satisfying the conditions

$$\theta_1(t_1)=\theta_0,\quad \theta_1(t_2)=\theta_2=v_1\theta_0/v_0;$$

$$p(t)=R\theta(t)/v(t)=\begin{cases} R\theta_0/v_1(t) & \text{for } t_0\le t\le t_1 \\ R\theta_1(t)/v_1 & \text{for } t_1\le t\le t_2 \\ p_0 & \text{for } t_2\le t\le t_3;\end{cases}$$

and

$$u(t)=\begin{cases} u_0 & \text{for } t_0\le t\le t_1 \\ u_0+c_v\theta_1(t) & \text{for } t_1\le t\le t_2 \\ u_0+c_v\theta_2+c_pp_0v_3(t)/R & \text{for } t_2\le t\le t_3.\end{cases}$$

It follows that

$$p\frac{dv}{dt}+\frac{du}{dt}=\begin{cases} (R\theta_0/v_1(t))v_1'(t)+0 & \text{for } t_0\le t\le t_1 \\ 0+c_v\theta_1'(t) & \text{for } t_1\le t\le t_2 \\ p_0v_3'(t)+c_pp_0v_3'(t)/R & \text{for } t_2\le t\le t_3.\end{cases}$$

Hence, by Eq. (1),

$$\frac{dq}{dt}=\begin{cases} R\theta_0v_1'(t)/v_1(t) & \text{for } t_0\le t\le t \\ c_v\theta_1'(t) & \text{for } t_1\le t\le t_2 \\ p_0(1+c_p/R)v_3'(t) & \text{for } t_2\le t\le t_3\end{cases}$$

and by integration

$$q(t)=\begin{cases} q(t_0)+R\theta_0 \ln[v_1(t)/v_0] & \text{for } t_0\le t\le t_1 \\ q(t_0)+R\theta_0 \ln(v_1/v_0) +c_v[\theta_1(t)-\theta_0] & \text{for } t_1\le t\le t_2 \\ q(t_0)+R\theta_0 \ln(v_1/v_0)+c_v(\theta_2-\theta_0) +p_0(1+c_p/R)[v_3(t)-v_1)] & \text{for } t_2\le t\le t_3.\end{cases}$$

In particular

$$q(t_3)=q(t_0)+R\theta_0 \ln(v_1/v_0)+c_v(\theta_2-\theta_0) +p_0(1+c_p/R)(v_3-v_1).$$

Transforming the last two terms we find

$$\Delta q=p_0\Delta v+c_p\Delta\theta+R\theta_0 \ln(v_1/v_0)-2p_0(v_1-v_0).$$

3. What Is and What Is Not a Function?

The last example also shows that if we merely know the net changes which the volume and the energy (or the temperature) of a gas undergo, we cannot even approximately guess how much heat has been transferred to the gas during the

process. For in the process Π_1, even if the net changes Δv, $\Delta\theta$, and Δu are as small as we please, Δq is appreciable. But, if Π_2 denotes the expansion of the volume of the gas under constant pressure from v_0 to v_3 (in the diagram, Π_2 would be represented by the straight segment AD), then we obtain

$$\Delta q = \int_{AD} (p_0 dv + c_p d\theta)$$
$$= p_0\Delta v + c_p\Delta\theta = p_0(1 + c_p/R)\Delta v.$$

Thus, for Π_2, if Δv is sufficiently small, Δq is as small as we please.

We thus see that Δq is not a function of Δv and $\Delta\theta$. Here, by "function" we mean the association of exactly one number with each element of some set. (This set is called the *domain* of the function.) If the elements of the set are numbers, pairs of numbers, triples of numbers, . . . we speak of a function of one, two, three, . . . variables. (Some authors call the functions in this sense *one-valued* or *single-valued* functions.) Moreover, we see that q is not a function of v and θ. For, if in Π_1 we set $v_3 = v_0$, then Δv, $\Delta\theta$, Δp are precisely zero, and yet

$$\Delta q = R\theta_0 \ln(v_1/v_0) - 2p_0(v_1 - v_0) < 0.$$

The beginning student of physics would easily grasp this situation, if in studying the concept of a function he had been shown examples of associations of numbers with numbers which do not fall under the definition. The temperature of a room is a function of the time, and the temperature of a man in the room is a function of the time; but the temperature of the man is not a function of the temperature of the room, since (at different moments) the man's temperature may be different although the room temperature is the same. Neither is the room temperature a function of the man's temperature. Even if, during a particular time interval, the man's temperature should be the same whenever the room temperature is the same, we still should not call the former a function of the latter since a little reflection would reveal the coincidental character of the particular situation.[2] Unfortunately, many mathematics books fail to present such examples.[3]

On the contrary, numerous books on calculus define, at the outset, a function as the association of one *or more* numbers with each element of a set (domain). According to this definition (of what sometimes is called a *many-valued* function), every measurable quantity varying in time is a function of every other such quantity; e.g., a man's temperature would indeed be a function of the room temperature. For this very reason, the general concept of a many-valued function is physically useless; moreover, it is also unimportant in the elements of mathematics. In calculus, we differentiate and integrate exclusively one-valued functions and even in the theory of complex functions we study only a very limited class of many-valued functions.

While thus the introduction of the general concept of a many-valued function does not serve any useful purpose, it creates in beginners the disposition to those misunderstandings which are particularly harmful in thermodynamics. Mathematicians ought to limit elementary discussions strictly to one-valued functions, but supplement the definition with examples not falling under the concept.

On the other hand, physicists ought to refrain from using such symbols[4] as $(\partial q/\partial v)_\theta$ which are meaningless except under the assumption that q is a function of v and θ—and this, as we have seen and as these physicists themselves point out, is not the case.

4. An Equation of State and the Energy

We shall now study a class of processes. Usually the processes of this class are called *quasi-static processes of a substance satisfying an equation of state*. Mathematically, these processes are characterized by a relation between volume, pressure, and temperature (which are said to

[2] Cf. the author's lecture notes "The concept of a function," edited by B. D. Fried (1948).

[3] Cooley, Gans, Kline, and Wahlert in their *Introduction to mathematics* (Houghton Mifflin, 1937), a book with many merits, claim (p. 262) that the marriages in New York are a function of the imports of Siam from the United States. The authors substantiate this claim by representing the marriages and imports as two functions of the time. By coincidence, whenever the imports differ so do the marriages. But even so, we should consider the situation as an example *not* falling under the concept of a function.

[4] H. Margenau and G. M. Murphy, *The mathematics of physics and chemistry* (Van Nostrand, 1943), p. 11.

characterize the *state* of the substance). The relation (equation of state) is assumed to hold for all quasi-static processes of the substance and to be of such a nature that every two of the quantities v, p, and θ determine the third. More precisely, we assume the existence of three functions

$$V(p, \theta), \quad P(v, \theta), \quad \Theta(v, p)$$

possessing continuous first partial derivatives which are so related that the following equalities hold

$$\begin{array}{ll} V(P(v, \theta), \theta) = v & \text{for every } v \text{ and } \theta, \\ V(p, \Theta(v, p)) = v & \text{for every } v \text{ and } p, \\ P(V(p, \theta), \theta) = p & \text{for every } p \text{ and } \theta, \\ P(v, \Theta(v, p)) = p & \text{for every } p \text{ and } v, \\ \Theta(V(p, \theta), p) = \theta & \text{for every } \theta \text{ and } p, \\ \Theta(v, P(v, \theta)) = \theta & \text{for every } \theta \text{ and } v. \end{array}$$

For every process of the class, at every moment t we have

$$v(t) = V(p(t), \theta(t)), \quad p(t) = P(v(t), \theta(t)), \quad \Theta(v(t), p(t)) = \theta(t)$$

and hence

$$v'(t) = \frac{\partial V}{\partial p} p'(t) + \frac{\partial V}{\partial \theta} \theta'(t) \tag{6}$$

as well as similar formulas for $p'(t)$ and $\theta'(t)$. Here $(\partial V/\partial p)$ and $(\partial V/\partial \theta)$ denote the functions of t obtained by substituting $p(t)$ and $\theta(t)$ into the partial derivatives of $V(p, \theta)$.

We shall furthermore assume that the energy is determined by volume, pressure, and temperature or, as we say, that u is a function of the state. In Sec. 2 we saw that, even for one and the same process, the quantity of heat need not be determined by volume, pressure, and temperature. In other words, q is not a function of the state.

In view of the equation of state, every function of the state, in particular u, is determined by any two of the three quantities v, p, and θ. This example illustrates the difference, mentioned in Paragraph (B) of the introduction, between the notations of the physicist and the mathematician.

(1) Since, in general, the way u depends upon v and p is totally different from the ways u depends upon v and θ, or upon p and θ, the mathematician, being exclusively interested in the way the function values depend upon the independent variables, would denote the three functions by three different symbols, say by

$$U_1(v, p), \quad U_2(v, \theta), \quad U_3(p, \theta)$$

with the understanding that we have

$$U_3(p, \theta) = U_2(V(p, \theta), \theta) \quad \text{for every } p \text{ and } \theta \tag{7_{math}}$$

$$U_3(p, \theta) = U_1(V(p, \theta), p) \quad \text{for every } p \text{ and } \theta \tag{8_{math}}$$

and similar formulas connecting any U_i to any U_j $(j \neq i)$. The mathematician would denote the six partial derivatives of the three functions by

$$\frac{\partial U_1}{\partial v}, \quad \frac{\partial U_1}{\partial p}, \quad \frac{\partial U_2}{\partial v}, \quad \frac{\partial U_2}{\partial \theta}, \quad \frac{\partial U_3}{\partial p}, \quad \frac{\partial U_3}{\partial \theta}.$$

By differentiating the relations (7_{math}), (8_{math}), etc., he would obtain the relations

$$\frac{\partial U_3}{\partial \theta} = \frac{\partial U_2}{\partial v} \frac{\partial V}{\partial \theta} + \frac{\partial U_2}{\partial \theta} \tag{9}$$

$$\frac{\partial U_3}{\partial p} = \frac{\partial U_2}{\partial v} \frac{\partial V}{\partial p}, \quad \text{etc.} \tag{10}$$

(2) The physicist, being primarily interested in the physical meaning of the function values, uses the same letter to denote the three functions and writes

$$u = u(v, p) = u(v, \theta) = u(p, \theta). \tag{7_{phys}}$$

In order to avoid ambiguities which would be disturbing from his as well as the mathematician's point of view, the physicist denotes the six partial derivatives of u by

$$\left(\frac{\partial u}{\partial v}\right)_p, \left(\frac{\partial u}{\partial p}\right)_v, \left(\frac{\partial u}{\partial v}\right)_\theta, \left(\frac{\partial u}{\partial \theta}\right)_v, \left(\frac{\partial u}{\partial p}\right)_\theta, \left(\frac{\partial u}{\partial \theta}\right)_p,$$

respectively. Accordingly he expresses the relations (9) and (10) in the form

$$\left(\frac{\partial u}{\partial \theta}\right)_p = \left(\frac{\partial u}{\partial v}\right)_\theta \left(\frac{\partial V}{\partial \theta}\right)_p + \left(\frac{\partial u}{\partial \theta}\right)_v \tag{9_{phys}}$$

and

$$\left(\frac{\partial u}{\partial p}\right)_\theta = \left(\frac{\partial u}{\partial v}\right)_\theta \left(\frac{\partial V}{\partial p}\right)_\theta, \tag{10_{phys}}$$

etc., where $(\partial V/\partial\theta)_p$ and $(\partial V/\partial p)_\theta$ could, without ambiguity, be replaced by $(\partial V/\partial\theta)$ and $(\partial V/\partial p)$, respectively.

The function of the time $u(t)$ is related to the three functions u of two variables as follows: At every moment t, we have

$$u(t)=u(v(t),\,p(t))=u(v(t),\,\theta(t))=u(p(t),\,\theta(t)) \quad (11_{\text{phys}})$$

$$u(t)=U_1(v(t),\,p(t))=U_2(v(t),\,\theta(t))=U_3(p(t),\,\theta(t)). \quad (11_{\text{math}})$$

Accordingly, we have, for instance,

$$\frac{du}{dt}=\left(\frac{\partial u}{\partial v}\right)_\theta\frac{dv}{dt}+\left(\frac{\partial u}{\partial \theta}\right)_v\frac{d\theta}{dt}, \quad (12_{\text{phys}})$$

$$u'(t)=\frac{\partial U_2}{\partial v}v'(t)+\frac{\partial U_2}{\partial \theta}\theta'(t), \quad (12_{\text{math}})$$

and two similar equalities for U_1 and U_3.

Once the coordination of the terminologies used in physics and mathematics is established, it is only fair for the physicist to insist on the merits of his own notation in the field of thermodynamics. For in developing this theme the physicist has to represent the energy as a function of many pairs of quantities besides (v, p), (v, θ), and (p, θ). The mathematician would have to introduce further symbols, such as u_4, u_5, $\cdots$ which would be unintuitive unless a definite order of all thermodynamical quantities were generally adopted.

It may be pointed out, furthermore, that in applying the concept of a function to geometry the mathematician himself often uses the notation of the physicist. He describes the equation of the motion of a point in the plane by $x=X(t)$, $y=Y(t)$ and denotes the curvature of the path at the moment t by $\kappa(t)$. Let $S(t)$ be the length of the path traversed between an initial moment t_0 and t. For given s, let $T(s)$ be the moment at which $S(T(s))=s$. Now most geometers write $\kappa(s)$ instead of $\kappa(T(s))$, although the way in which the curvature depends upon s is, in general, totally different from the way in which the curvature depends upon t. However, it seems doubtful that one could develop a consistent, purely logical syntax of this physical-geometrical functional notation, except with regard to partial derivatives.

5. The Specific Heat

Traditionally, the first application of the first law is made in computing the specific heat. For a given differentiable process Π we call $(q'(t)/\theta'(t))$ the specific heat at the moment t. We have

$$\frac{q'(t)}{\theta'(t)}\sim_{\Delta t}\frac{q(t+\Delta t)-q(t)}{\theta(t+\Delta t)-\theta(t)}.$$

From the first law, Eq. (1), we obtain

$$\frac{q'(t)}{\theta'(t)}=\frac{p(t)v'(t)+u'(t)}{\theta'(t)}. \quad (13)$$

(*a*) *Constant volume.*—First we consider a process, Π_{v_0} during which the volume is constant, $v(t)=v_0$, and thus $v'(t)=0$. For every such process Π_{v_0} belonging to the class studied in Sec. 4, from Eqs. (13) and (12) in view of $v'(t)=0$ we obtain

$$c_v(t)=\left(\frac{q'(t)}{\theta'(t)}\right)_v=\begin{cases}(\partial u/\partial\theta)_v & \text{in the physical notation,}\\ (\partial U_2/\partial\theta)(v_0,\,\theta(t)) & \text{in the mathematical notation.}\end{cases}$$

Here the last term indicates that in $(\partial U_2/\partial\theta)$ we substitute v_0 for v, and $\theta(t)$ for θ.

We see that, if for a process Π_{v_0} at two different moments the temperature has the same value (the volume has automatically the same value v_0), then at these moments the specific heat will be the same. But much more can be said. Even if the substance with the energy function $u(v, \theta)$ undergoes different processes, Π_{v_0} and $\Pi_{v_0}^*$, with constant volume v_0, then whenever the temperatures are equal, so are the specific heats—in other words, the specific heat for constant volume is a function of the state.

We can also express this fact by saying that there exists a function $c(v, \theta)$ such that, if a substance with the energy function $u(v, \theta)$ undergoes any process Π_{v_0}, then at every moment t the specific heat is

$$c_{v0}(t)=c(v_0,\,\theta(t)).$$

(*b*) *Constant pressure.*—Let Π_{p_0} denote a process of our substance for which $p(t)=p_0$. Then

$p'(t)=0$ and from Eqs. (13), (6), and the analog of (12) for $U_3=u(p,\theta)$ it follows that

$$c_{p0}(t)=\frac{q'(t)}{\theta'(t)}=p_0\frac{\partial V}{\partial\theta}(p_0,\theta(t))+\frac{\partial U_3}{\partial\theta}(p_0,\theta(t))$$

$$=p_0\left(\frac{\partial V}{\partial\theta}\right)_p+\left(\frac{\partial u}{\partial\theta}\right)_p.$$

Using Eq. (9_{phys}) we obtain

$$c_{p0}(t)=p_0\left(\frac{\partial V}{\partial\theta}\right)_p+\left(\frac{\partial u}{\partial v}\right)_\theta\left(\frac{\partial V}{\partial\theta}\right)_p+\left(\frac{\partial u}{\partial\theta}\right)_v.$$

We see also that c_{p0} is a function of the state. Again, conditions can be formulated under which c_p does not change during the process, and assumes the same value for all processes Π_p within a certain range. Clearly,

$$c_p-c_v=[p+(\partial u/\partial v)_\theta](\partial V/\partial\theta)_p.$$

6. An Integral Form of the First Law

The form Eq. (1) of the first law is confined to differentiable processes for which the volume, the energy, and the quantity of heat are differentiable functions of the time. We shall now present an integral form of the first law which does not presuppose differentiability of any of the functions, and thus is applicable to a much wider class of processes.

We shall base this formulation on the concept of the Stieltjes integral of a function $f(x)$ with respect to a function $g(x)$ in an interval, $a\leq x\leq b$, denoted by

$$\int_a^b f(x)dg(x).$$

Due to its fundamental role in pure as well as applied mathematics, this concept has lately been included in several elementary textbooks.[5] Its main applications are to moments in mechanics and statistics. As we shall see, Stieltjes integrals are also useful in thermodynamics.

The idea of a Stieltjes integral is so simple that it can easily be presented to a beginning student of theoretical physics. In order to define $\int_a^b f(x)dg(x)$ we divide the interval $a\leq x\leq b$ into smaller intervals

$$a=x_0<x_1<\cdots<x_k<x_{k+1}<\cdots<x_{n-1}<x_n=b. \quad (14)$$

Let λ denote the length of the longest of these intervals, that is, the largest of the numbers $x_{k+1}-x_k$. In each of the small intervals we choose a number $x_k\leq x_k^*\leq x_{k+1}$. Then we form the so-called Stieltjes sum

$$\sum_{k=0}^{n-1} f(x_k^*)[g(x_{k+1})-g(x_k)]$$

which is an approximation to $\int_a^b f(x)dg(x)$. In order to obtain a better approximation we form Stieltjes sums for subdivisions of the interval $a\leq x\leq b$ which are finer, that is, for which λ is smaller. The integral $\int_a^b f(x)dg(x)$ is defined as the limit, whenever it exists, of the Stieltjes sums as λ approaches zero. In the special case that $g(x)=x$ the integral is the ordinary integral $\int_a^b f(x)dx$. In the special case that $f(x)=1$ for $a\leq x\leq b$ we readily see that $\int_a^b 1dg(x)=g(b)-g(a)$ for every function $g(x)$. If $g(x)$ is differentiable, one can easily prove that

$$\int_a^b f(x)dg(x)=\int_a^b f(x)g'(x)dx.$$

But, the Stieltjes integral also exists for non-differentiable functions $g(x)$ and even for some discontinuous functions $g(x)$; for instance, it exists for functions having only a finite number of maxima and minima in the interval $a\leq x\leq b$, provided that $f(x)$ is not too discontinuous.

The function $g(x)$ is said to be of *bounded variation* in the interval $a\leq x\leq b$, if there exists a finite number V such that for every subdivision (Eq. (14)) of the interval we have

$$|g(x_1)-g(x_0)|+\cdots+|g(x_{k+1})-g(x_k)|+\cdots+|g(x_n)-g(x_{n-1})|\leq V.$$

It can be shown that every continuous function having only a finite number of maxima and minima in an interval is of bounded variation in this interval. The converse of this theorem is not true. The continuous function

$$g(x)=\begin{cases}x^2\sin(1/x) & \text{for } 0<x\leq 1\\ 0 & \text{for } x=0\end{cases}$$

[5] D. V. Widder, *Advanced calculus* (Prentice-Hall, 1947), Chap. V.

can be proved to be of bounded variation in the interval $0 \leq x \leq 1$, although it has infinitely many maxima and minima in the interval.

It can be shown that $\int_a^b f(x)dg(x)$ exists whenever $f(x)$ is continuous and $g(x)$ of bounded variation in the interval $a \leq x \leq b$. The integral form of the first law reads

$$q(t_1) - q(t_0) = u(t_1) - u(t_0) + \int_{t_0}^{t_1} p(t)dv(t). \quad (15)$$

If $v(t)$ is differentiable, then the last integral is $= \int_{t_0}^{t_1} p(t)v'(t)dt$. We can divide equality (15) by $t_1 - t_0$. If also $q(t)$ and $u(t)$ are differentiable, and we let t_1 approach t_0, then we obtain the differential form, Eq. (1), of the first law for $t = t_0$. But from what was said about Stieltjes integrals it follows that the integral form of the first law is also applicable to processes which are not differentiable, even to processes in which the volume undergoes discontinuous changes, provided that at the moments of these explosions the pressure is continuous.

7. Expansion and Contraction

The remark contained in this section is due to Mr. Leo A. Schmidt. We call a process Π_m an *expansion*, if $a \leq t_1 \leq t_2 \leq b$ implies $v(t_1) \leq v(t_2)$ and $v(a) < v(b)$; a *contraction*, if $a \leq t_1 < t_2 \leq b$ implies $v(t_1) \geq v(t_2)$ and $v(a) > v(b)$. The subscript in Π_m indicates that, in both cases, $v(t)$ is what is called a *monotonic* function. If $g(x)$ is monotonic for $a \leq x \leq b$, then the Stieltjes integral $\int_a^b f(x)dg(x)$ satisfies the following mean value theorem: There exists a number x^* between a and b such that

$$\int_a^b f(x)dg(x) = f(x^*)[g(b) - g(a)].$$

From the integral form (15) of the first law it thus follows for Π_m that

$$q(t_1) - q(t_0) = p(t^*)[v(t_1) - v(t_0)] + u(t_1) - u(t_0)$$

for some t^* between t_0 and t_1, or briefly, that

$$\Delta q = p^* \Delta v + \Delta u.$$

If we set $p(t_0) = p$, it follows that

$$\Delta q - p\Delta v - \Delta u = (p^* - p)\Delta v.$$

Now, if we denote by $\Delta p_{\max}$ the maximum change that the pressure undergoes during the process Π_m then we have

$$|p^* - p| \leq \Delta p_{\max}$$

and

$$\left| \frac{\Delta q - p\Delta v - \Delta u}{\Delta v} \right| \leq \Delta p_{\max}.$$

In words: *In an expansion or contraction, if the change in pressure is sufficiently small, the difference between Δq and $p\Delta v + \Delta u$ is as small in comparison with Δv as we please.*

8. What is a Differential?

We shall discuss the case of three variables from which the transition to two or one, as well as to four or more variables, is easy.

Let f be a function of three variables which in a domain D admits three continuous first partial derivatives. For every triple (x, y, z) belonging to the domain D, and every triple of numbers $(\Delta x, \Delta y, \Delta z)$, we form the number

$$\Lambda_f(x, y, z; \Delta x, \Delta y, \Delta z) = \frac{\partial f}{\partial x}(x, y, z)\Delta x + \frac{\partial f}{\partial y}(x, y, z)\Delta y + \frac{\partial f}{\partial z}(x, y, z)\Delta z.$$

If we set $\Delta r = [(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2]^{\frac{1}{2}}$, then, as is taught in the elements of calculus,

$$\Lambda_f(x, y, z; \Delta x, \Delta y, \Delta z) \sim_{\Delta r} f(x + \Delta x, y + \Delta y, z + \Delta z) - f(x, y, z).$$

That is, for small Δr the difference between Λ_f and Δf is not only small—this is obvious because for small Δr both Λ_f and Δf themselves are small—but small in comparison with Δr. In other words, by choosing Δr sufficiently small we can make

$$\left| \frac{\Delta f - \Lambda_f}{\Delta r} \right| \quad (16)$$

as small as we please. [The statement mentioned in Sec. 1 about a function $f(t)$ is the analog for functions of one variable of our last statement about $f(x, y, z)$.]

For every triple (x, y, z) belonging to D the following *difference function* of f at (x, y, z), $\Delta f = f(x + \Delta x, y + \Delta y, z + \Delta z) - f(x, y, z)$, is a func-

tion (and, in general, a complicated function) of Δx, Δy, Δz, while $\Lambda_f(x, y, z; \Delta x, \Delta y, \Delta z)$, which we call the *differential* of f at (x, y, z), is a linear function of Δx, Δy, Δz.

The significance of the above result lies in the fact that, for sufficiently small Δr, the inaccuracy we incur by replacing the complicated difference function by the linear differential is as small as we please in comparison with Δr.

Traditionally, the last fact is symbolized as follows

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz. \tag{17}$$

In presenting Eq. (17) to the beginning mathematician and physicist one cannot emphasize too strongly that this formula is neither more nor less than a symbol for the possibility of approximating Δf by Λ_f in the way explained in the preceding paragraph. It is the conviction of the author that it would be best not to present to beginners the symbolic expression Eq. (17) at all.

Now let us assume that in a domain D of the (x, y, z)-space three continuous functions, $P(x, y, z)$, $Q(x, y, z)$, and $R(x, y, z)$, are given. For every triple (x, y, z) belonging to D and every triple of numbers $(\Delta x, \Delta y, \Delta z)$ we may form the number

$$\Lambda_{P,Q,R}(x, y, z; \Delta x, \Delta y, \Delta z) = P(x, y, z)\Delta x + Q(x, y, z)\Delta y + R(x, y, z)\Delta z.$$

In this way for given P, Q, and R, we have defined a function of six variables. We can also say: given P, Q, and R, for every triple (x, y, z) belonging to D, we have defined a linear function of $(\Delta x, \Delta y, \Delta z)$. If, in particular, P, Q, and R are the partial derivatives of a function $f(x, y, z)$

$$P = \partial f/\partial x, \quad Q = \partial f/\partial y, \quad R = \partial f/\partial z,$$

then $\Lambda_{P,Q,R}(x, y, z; \Delta x, \Delta y, \Delta z)$ is what we have denoted by $\Lambda_f(x, y, z; \Delta x, \Delta y, \Delta z)$ and called the differential of f. For this reason, if P, Q, and R are any three given continuous functions, $\Lambda_{P,Q,R}$ has received the rather unfortunate name of a (general) *differential* while the differential Λ_f of a function is called a *complete* or *exact* differential.

The traditional symbols

$$P(x, y, z)dx + Q(x, y, z)dy + R(x, y, z)dz$$

for the (general) differential, shaped after the symbols on the right side of Eq. (17), are, as we shall see, even more unfortunate. At this place we just emphasize that in our definition of $\Lambda_{P,Q,R}$ the numbers Δx, Δy, Δz are in no way restricted and, in particular, need not be small.

Summarizing we can say: Λ_{f_x, f_y, f_z} is the differential of f. The differential $\Lambda_{P,Q,R}$, for three given functions P, Q, and R with the domain D, is the association of the linear function $\Lambda_{P,Q,R}(x, y, z; \Delta x, \Delta y, \Delta z)$ of $(\Delta x, \Delta y, \Delta z)$ with every (x, y, z) belonging to D.

For instance, if we have

$$P(x, y, z) = 3x^2 + yz + 2, \quad Q(x, y, z) = 0, \quad R(x, y, z) = 4x$$

for the domain of all triples (x, y, z), then with $(0, 0, 0)$, $(1, 1, 1)$, and $(3, 2, 1)$ we associate the linear functions

$$\Lambda_{P,Q,R}(0, 0, 0; \Delta x, \Delta y, \Delta z) = 2\Delta x,$$
$$\Lambda_{P,Q,R}(1, 1, 1; \Delta x, \Delta y, \Delta z) = 6\Delta x + 4\Delta z,$$
$$\Lambda_{P,Q,R}(3, 2, 1; \Delta x, \Delta y, \Delta z) = 31\Delta x + 12\Delta z.$$

The concept of a differential is thus exceedingly simple. But it is of that kind of simplicity and abstractness which baffles beginners. The usual symptom of their lack of understanding is their response to the above definition of $\Lambda_{P,Q,R}$ with the question "But what *is* a differential?" They will not ask this question with regard to the differential of a function f. For in the mind of the beginner, Λ_f gets its real meaning from the fact that it is approximately equal to Δf, the difference between values of f. This point of view is, of course, incorrect. *By definition*, Λ_f is a linear function of $(\Delta x, \Delta y, \Delta z)$ for every (x, y, z), precisely as is $\Lambda_{P,Q,R}$; and it is a *theorem* that, if Λ_f is so defined, then Λ_f and Δf differ by a quantity which is as small as we please in comparison with $\Delta r = [(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2]^{\frac{1}{2}}$ provided that Δr is sufficiently small. But the tie between Λ_f and Δf is what really interests the beginner. No similar tie exists between $\Lambda_{P,Q,R}$ and more familiar concepts. The result is that $\Lambda_{P,Q,R}$ is not understood by beginners with the exception of a small minority.

The traditional expression, $Pdx + Qdy + Rdz$, for the differential adds to the confusion. For in the symbolic equality Eq. (17) we write dx, dy, and dz in order to indicate that the inaccuracy incurred by replacing Δf by Λ_f gets smaller and

smaller in comparison with $|\Delta x|$, $|\Delta y|$, and $|\Delta z|$ as these quantities get sufficiently small. Hence the expression $Pdx+Qdy+Rdz$ suggests that some inaccuracy gets smaller or, at any rate, that something happens to $P\Delta x+Q\Delta y+R\Delta z$ when $|\Delta x|$, $|\Delta y|$, and $|\Delta z|$ get sufficiently small. As a matter of fact, however, in general nothing particular happens under these circumstances.

The conclusion which the author draws from these facts is *that physicists in presenting the elements of the theory should refrain from referring to the poorly understood concept of a general differential altogether.*

How such references can easily be avoided in presenting the first law has been shown in Sec. 1 where we introduced the law as the relation (1) between time derivatives, and in Sec. 6 where we presented the integral form Eq. (15).

After the beginner has really understood the first law in these forms which he is prepared to understand, the physicist may, for historical reasons, wish to present the differential form Eq. (3). In doing so he ought to emphasize the remark made in Sec. 2 that the differentials are meant with respect to the time, and he should present examples, such as those of Sec. 2, which forestall misinterpretations of the differentials in other directions. For

$$p\Delta v+\Delta u \quad \text{and} \quad [p+(\partial u/\partial v)]\Delta v+(\partial u/\partial\theta)\Delta\theta$$

are general differentials in two variables v, u and v, θ, respectively. For instance, the latter differential can be written in the form

$$A(v,\theta)\Delta v+B(v,\theta)\Delta\theta,$$

where

$$A(v,\theta)=P(v,\theta)+(\partial u/\partial v)(v,\theta)$$

and $B(v,\theta)=(\partial u/\partial\theta)(v,\theta)$, $P(v,\theta)$ representing the pressure as a function of volume and temperature. In general, the differential $A(v,\theta)\Delta v+B(v,\theta)\Delta\theta$ is not the differential of a function $f(v,\theta)$. In their futile attempts to connect this differential with more familiar concepts beginners are misled by the unfortunate form Eq. (3) of the first law into the misinterpretations discussed in Sec. 3 which cannot be prevented by simply replacing the symbol dq by $đq$, δq, or Q.

9. A General Scheme for the Elimination of Differentials

More easily comprehensible to the beginner than the differential $\Lambda_{P,Q,R}(x,y,z;\Delta x,\Delta y,\Delta z)$ is the simple idea of the triple of functions $(P(x,y,z),\ Q(x,y,z),\ R(x,y,z))$ because it may be interpreted as the association of a vector with each point. With the point (x,y,z) we associate the vector with the components $P(x,y,z)$, $Q(x,y,z)$, and $R(x,y,z)$.

The possibility of replacing the idea of the differential $\Lambda_{P,Q,R}$ by that of the vector field (P,Q,R) is most obvious in mechanics. A force field in the three-dimensional space is given by the association of a vector $(P(x,y,z),\ Q(x,y,z),\ R(x,y,z))$ with each point (x,y,z) of a domain.

The differential with the classical symbol

$$P(x,y,z)dx+Q(x,y,z)dy+R(x,y,z)dz \quad (18)$$

means that for every point (x,y,z) we define a linear function of $(\Delta x,\Delta y,\Delta z)$; that is, that, for every point (x,y,z), with every vector from (x,y,z) to $(x+\Delta x,\ y+\Delta y,\ z+\Delta z)$ we associate a number, namely, the scalar product of this vector with the vector $(P(x,y,z),\ Q(x,y,z),\ R(x,y,z))$. The physical meaning of this scalar product is the work we should do by moving a unit mass from (x,y,z) to $(x+\Delta x,\ y+\Delta y,\ z+\Delta z)$ along the straight line, if the force vector everywhere along this path were identical with the force vector at the initial point (x,y,z) which, in general, is, of course, not the case.

Now, in spite of the fact that in some respects the concept of work is more basic than that of force, one will probably hesitate to introduce the beginner to the theory of force fields by taking the differential Eq. (18) as the starting point. One will much rather begin with the force field and then define the work W done by moving a unit mass along the path $x=x(t)$, $y=y(t)$, $z=z(t)$, $(t_0\leq t\leq t)$ as the Stieltjes integral

$$\int_{t_0}^{t_1}[P(x(t),y(t),z(t))dx(t) \\ +Q(x(t),y(t),z(t))dy(t) \\ +R(x(t),y(t),z(t))dz(t)]$$

corresponding to the integral form Eq. (15) of the first law. For differentiable motions, we may

relate the time derivatives of W and the components of the velocity,

$$W'(t)=P(x(t), y(t), z(t))x'(t) \\ +Q(x(t), y(t), z(t))y'(t) \\ +R(x(t), y(t), z(t))z'(t)$$

corresponding to the form Eq. (1) of the first law.

The general procedure for eliminating differentials $\Lambda_{P,Q,R}$ is to concentrate on the triple of functions $(P(x, y, z), Q(x, y, z), R(x, y, z))$. For instance, we shall not say that the differential $Pdx+Qdy+Rdz$ is complete or exact, or, in other words, that there exists a function f such that $df=Pdx+Qdy+Rdz$. Instead we shall say that the triple of functions (P, Q, R) is *exact* if, and only if, there exists a function f such that

$$P=\partial f/\partial x, \quad Q=\partial f/\partial y, \quad R=\partial f/\partial z.$$

Similarly, we shall say, the triple (P, Q, R) is *integrable* if, and only if, there exists a function $s(x, y, z)$ such that the triple (sP, sQ, sR) is exact. The beginner can be expected to know that every pair of functions of two variables $(P(x, y), Q(x, y))$, but not every triple of functions of three variables, is integrable.

10. The Second Law

The second law, in its traditional form, is the conjunction of various, more or less independent, statements. We shall mention the three main mathematical consequences of the second law.

(*a*) *The existence of the entropy.*—There exists a function of the state, $S(v, \theta)$, for which

$$\frac{\partial S}{\partial v}:\frac{\partial S}{\partial \theta}=\left[P(v, \theta)+\left(\frac{\partial u}{\partial v}\right)_\theta\right]:\left(\frac{\partial u}{\partial \theta}\right)_v.$$

In the terminology of the preceding section, statement (*a*) contends that the pair of functions, $P(v, \theta)+(\partial u/\partial v)_\theta$, $(\partial u/\partial \theta)_v$ is integrable or, in the classical terminology, that the differential

$$[P(v, \theta)+(\partial u/\partial v)_\theta]dv+(\partial u/\partial \theta)_v d\theta$$

is integrable. As we mentioned at the end of the preceding section, every pair of two functions of two variables and every differential in two variables is integrable. Hence, in the case of processes which can be characterized by two functions $v(t)$ and $\theta(t)$, statement (*a*) is provable and thus does not impose any limitations on the course of nature.

The situation changes, if we study more complicated processes, for instance, processes in which each state is characterized by four numbers v_1, θ_1, v_2, and θ_2 describing volume and temperature of two substances. In this case, the corresponding quartet of functions or, in the classical terminology, the differential

$$\left[P_1(v_1, \theta_1)+\left(\frac{\partial u_1}{\partial v}\right)_{\theta_1}\right]dv_1+\left(\frac{\partial u_1}{\partial \theta_1}\right)_{v_1}d\theta_1 \\ +\left[P_2(v_2, \theta_2)+\left(\frac{\partial u_2}{\partial v_2}\right)_{\theta_2}\right]dv_2+\left(\frac{\partial u_2}{\partial \theta_2}\right)_{v_2}d\theta_2 \quad (19)$$

need not be and, in general, is not integrable. For instance, one readily proves that the differential

$$\frac{R\theta_1}{v_1}dv_1+c_1d\theta_1+\frac{R\theta_2}{v_2}dv_2+c_2d\theta_2 \quad (20)$$

corresponding to two ideal gases, is not integrable. For the assumption that there exists a function $S(v_1, \theta_1, v_2, \theta_2)$ such that

$$\frac{\partial S}{\partial v_1}:\frac{\partial S}{\partial \theta_1}:\frac{\partial S}{\partial v_2}:\frac{\partial S}{\partial \theta_2}=\frac{R\theta_1}{v_1}:c_1:\frac{R\theta_2}{v_2}:c_2$$

leads to a contradiction. However, in the case of two substances, the analog of statement (*a*) postulates the integrability only if the two substances have the same temperature, that is, if $\theta_1=\theta_2=\theta$. The differential Eq. (20) reads in this case

$$\frac{R\theta}{v_1}dv_1+\frac{R\theta}{v_2}dv_2+(c_1+c_2)d\theta \quad (21)$$

and Eq. (21) is indeed integrable. For, let S be the function

$$S(v_1, v_2, \theta)=R\log v_1+R\log v_2+(c_1+c_2)\log\theta. \quad (22)$$

Then we have

$$\frac{\partial S}{\partial v_1}=\frac{R}{v_1}, \quad \frac{\partial S}{\partial v_2}=\frac{R}{v_2}, \quad \frac{\partial S}{\partial \theta}=\frac{c_1+c_2}{\theta}$$

and, thus, obviously

$$\frac{\partial S}{\partial v_1}:\frac{\partial S}{\partial v_2}:\frac{\partial S}{\partial \theta}=\frac{R\theta}{v_1}:\frac{R\theta}{v_2}:c_1+c_2.$$

Hence the differential Eq. (21) is compatible with proposition (*a*).

One easily sees that, for instance, the differential $\theta dv_1+dv_2+d\theta$ is incompatible with proposition (*a*). Hence proposition (*a*) precludes the possibility of processes corresponding to the above differential.

We finally mention the obvious fact that for every pair of functions of two variables there exist *many* functions whose partial derivatives are proportional to the functions of the pair. For instance, if

$$\frac{\partial S}{\partial v}:\frac{\partial S}{\partial \theta}=P(v_1\theta)+\left(\frac{\partial u}{\partial v}\right)_\theta:\left(\frac{\partial u}{\partial \theta}\right)_v,$$

then, for every function f of one variable if $T(v,\theta)=f(S(v,\theta))$, we have

$$\frac{\partial T}{\partial v}:\frac{\partial T}{\partial \theta}=\frac{\partial S}{\partial v}:\frac{\partial S}{\partial \theta}.$$

Similarly, if a triple of functions or a differential in three variables, such as Eq. (21), is integrable, there exist many functions S satisfying the required conditions. One of them is selected according to physical principles and called the *entropy*.

(*b*) *The nature of the integrating denominator.*—The ratio

$$(\partial S/\partial v):[P(v,\theta)+(\partial u/\partial v)_\theta]$$

which, by statement (*a*), is equal to the ratio $(\partial S/\partial\theta):(\partial u/\partial\theta)_v$, is independent of v. If we denote this ratio by $1/N(\theta)$, then $N(\theta)$ is called an *integrating denominator* of the pair of functions $P(v,\theta)+(\partial u/\partial v)_\theta$, $(\partial u/\partial\theta)_v$ or of the corresponding differential.

While the exactness of this pair of functions of two variables, and hence the existence of *some* integrating denominator, is provable (as we emphasized in discussing statement (*a*), the existence of a *particular* denominator, namely, of a denominator which is independent of v (asserted in statement (*b*)), is a law of nature. This law precludes, for instance, the possibility that

$$P(v,\theta)+(\partial u/\partial v)_\theta=v \text{ and } (\partial u/\partial\theta)_v=e^v.$$

Statement (*b*) is even more restrictive with regard to processes in which two or more substances (at the same temperature θ) are involved. But we see that the differential Eq. (21) satisfies the condition of statement (*b*) since $N(\theta)=\theta$ is an integrating denominator.

From statements (*a*) and (*b*) in conjunction with the first law we obtain

$$\frac{\partial S}{\partial v}(v(t),\theta(t))v'(t)+\frac{\partial S}{\partial \theta}(v(t),\theta(t))\theta'(t)$$
$$=\frac{P(v(t),\theta(t))+(\partial u/\partial v)_\theta}{N(\theta(t))}v'(t)+\frac{(\partial u/\partial\theta)_v\theta'(t)}{N(\theta(t))}=\frac{q'(t)}{N(\theta(t))}.$$

If for every moment t we set

$$s(t)=S(v(t),\theta(t)),$$

we have

$$s'(t)=q'(t)/N(\theta(t))$$

and

$$s(t_1)-s(t_0)=\int_{t_0}^{t_1}\frac{dq(t)}{N(\theta(t))}.$$

The last Stieltjes integral might be used as a definition of $s(t)$ even in cases when $q(t)$ is a not differentiable function.

As a corollary, we deduce: If Π_{q_0} is an adiabatic quasistatic process for which $q(t)=q_0$ during the period $t_0\leq t\leq t_1$, then $s(t_1)-s(t_0)=0$. In words: *During an adiabatic quasi-static process, the entropy remains constant.* This statement is also valid for processes in which two or more substances are involved.

(*c*) *The monotony of the entropy.* During an adiabatic generalized process Γ the entropy never decreases.

Since we have just seen that during an adiabatic quasi-static process the entropy remains constant we first illustrate statement (*c*) by a nonquasi-static process during which the entropy increases; secondly, we mention a process during which the entropy decreases and which, consequently, by statement (*c*) is impossible in nature.

We let an ideal gas undergo the nonquasi-static process mentioned in Sec. 1 in such a way that both containers are adiabatically isolated from the rest of the world. Then we have

$$v(t_0)=v_0,\quad v(t_1)=2v_0,$$
$$q(t)=q_0 \quad\text{for}\quad t_0\leq t\leq t_1.$$

Experience shows that the temperature also remains constant during this process

$$\theta(t)=\theta_0 \quad \text{for} \quad t_0 \leq t \leq t_1.$$

In analogy to formula (22), for the entropy of an ideal gas characterized by its volume v and its temperature θ, we obtain expression $S(v, \theta) = R \log v + c \log \theta$. If $s(t)$ denotes the entropy of the gas at the moment t, we thus have

$$s(t_0)=S(v_0, \theta_0), \quad s(t_1)=S(2v_0, \theta_0)$$
$$s(t_0)=S(v_0, \theta_0)=R \log v_0 + c \log \theta_0$$
$$s(t_1)=S(2v_0, \theta_0)=R \log(2v_0) + c \log \theta_0,$$

thus

$$s(t_0) < s(t_1).$$

A nonquasi-static process precluded by statement (*c*) is the reverse of the above expansion, that is, a spontaneous adiabatic contraction of a gas to half of its initial volume. Since in this case

$$s(t_0)=S(2v_0, \theta), \quad s(t_1)=S(v_0, \theta_0)$$

we should have

$$s(t_0) > s(t_1)$$

which is impossible by statement (*c*).

Appendix. Which Processes are Thermodynamically Equal?

Let M be a continuous *motion* of a particle in the (x, y)-plane. We describe M by two continuous functions $X(t)$ and $Y(t)$, both defined for the interval $t_0 \leq t \leq t_1$ where, for every t of this interval, $X(t)$ denotes the abscissa, $Y(t)$ the ordinate of the particle at the moment t. By the *trace* of the motion M we shall mean the set (or locus) of all points (x, y) traversed by the moving point; that is, (x, y) belongs to the trace of M if, and only if, there exists at least one number t such that $t_0 \leq t \leq t_1$ and $x = X(t)$ and $y = Y(t)$.

Now let M^* be a continuous motion of a particle in the same plane, described by the two functions $X^*(u)$ and $Y^*(u)$ both defined for $u_0 \leq u \leq u_1$. The motions M and M^* are *identical* if, and only if, they occur during the same time interval and, at every moment, the positions of the two particles are identical; that is, if, and only if, $t_0 = u_0$, $t_1 = u_1$ and $X(s) = X^*(s)$, $Y(s) = Y^*(s)$ for every s of the interval.

Besides this strict identity, we are also interested in weaker relations between M and M^*. For instance, we might call M and M^* *kinematically equal* if they differ only inasmuch as they occur at different time intervals but become identical if one of the clocks used in describing the motions, say the second, is advanced by c time units; that is, if and only if $t_0 = u_0 + c$, $t_1 = u_1 + c$, $X^*(u) = X(u+c)$, and $Y^*(u) = Y(u+c)$ for every u between u_0 and u_1.

We might call M and M^* *kinematically similar* if they become identical when the second clock, at the moment 0, is accelerated in the ratio $b:1$ and then advanced by c time units; that is to say, if $t_0 = bu_0 + c$, $t_1 = bu_1 + c$, $X^*(u) = X(bu+c)$, and $Y^*(u) = Y(bu+c)$ for every u between u_0 and u_1.

When shall we say that M and M^* are *geometrically equal* or, as it is also expressed, that M and M^* determine *the same path*? A *necessary* condition is, of course, that M and M^* have the same trace. But this condition is by no means sufficient. For the mere trace of a motion does not, for instance, determine the length of the path traversed, and certainly this length is geometrically relevant. For example, if we merely know that the trace of a motion is the set of all points on the circle $x^2 + y^2 = r^2$, then we know that the length of the path is at least $2\pi r$. But it is $2n\pi r$, if the moving particle traverses the circle n times according to the equations $x = r \cos nt$, $y = r \sin nt$ for $0 \leq t \leq 2\pi$. In fact, any number $\geq 2\pi r$ is the length of the path of some motion having the above circle as its trace.

A *sufficient* condition for geometric identity is kinematic equality or similarity. But even similarity is by no means necessary. If, for instance, M_1 is described by

$$x = r \cos\sqrt{t}, \quad y = r \sin\sqrt{t} \quad \text{for} \quad 0 \leq t \leq 4\pi^2$$

and M_1^* is described by

$$x = r \cos u, \quad y = r \sin u \quad \text{for} \quad 0 \leq u \leq 2\pi,$$

then M_1 and M_1^* are not kinematically similar although they are equal in every geometrical respect.

These remarks show that the geometric equality of motions or the equality of their paths is less than their kinematic similarity and more than the mere identity of their traces.[6]

In the last example as in the case of kinematically similar motions we see that the time intervals $0 \leq t \leq 4\pi^2$ and $0 \leq u \leq 2\pi$ during which the motions take place, are monotonically related in such a way that in related moments the positions of the particles are identical. The relation is established by the function $t = u^2$ which amounts to a (nonlinear) distortion of the time scale satisfying the condition of monotony. For if the moment u' precedes the moment u'', then the moment $t' = u'^2$ precedes the moment $t'' = u''^2$.

Next we consider the following two motions along the X axis: M_2 described by the functions

$$X(t) = \begin{cases} 2t & \text{for } 0 \leq t \leq 1/2 \\ 1 & \text{for } 1/2 \leq t \leq 1 \end{cases}, \quad Y(t) = 0 \quad \text{for} \quad 0 \leq t \leq 1,$$

[6] A path can thus be defined either as *a class of motions* (namely, as the class of all motions which are geometrically equal to any motion of the class) or as *a trace with some additional information* (namely, information from which we can infer the way the trace has been traversed). This intrinsic definition of a path has been developed in the author's note "Définition intrinsèque de la notion de chemin," *Comptes Rendus* **221**, 739 (1945). In what follows, we apply the first definition to thermodynamical processes. But also our intrinsic definition can be applied to thermodynamics.

and M_2^* described by the functions

$$X(u)=\begin{cases}\frac{3}{2}u & \text{for } 0\leq u\leq 1/3\\ 1/2 & \text{for } \frac{1}{3}\leq u\leq 2/3,\\ \frac{3}{2}u-\frac{1}{2} & \text{for } \frac{2}{3}\leq u\leq 1\end{cases}\qquad Y(u)=0 \text{ for } 0\leq u\leq 1.$$

The first particle moves from $(0, 0)$ to $(1, 0)$ and rests there for one-half time unit; the second moves from $(0, 0)$ to $(\frac{1}{2}, 0)$, rests there for one-third time unit, and then moves to $(1, 0)$. We call $\frac{1}{2}\leq t\leq 1$ and $\frac{1}{3}\leq u\leq\frac{2}{3}$ *rest intervals* of the two motions. In the case of M_2 and M_2^* it would be impossible to establish a one-to-one relation or a strictly monotonic relation between the time intervals in such a way that at related moments the positions of the moving particles are identical. Yet M_2 and M_2^* are equal in every geometric respect. For the differences regarding the speed or the rest intervals are kinematic rather than geometric. The motions M_2 and M_2^* become identical if the clocks by means of which we describe the motions are changed in the following way. We let the first clock stop between $t=\frac{1}{2}$ and $t=1$ so that the entire motion takes place between $t=0$ and $t=\frac{1}{2}$. The second clock is decelerated in the ratio 3:4 between $u=0$ and $u=\frac{1}{3}$ so that the first part of M_2^* takes place between 0 and $\frac{1}{4}$; between $u=\frac{1}{3}$ and $u=\frac{2}{3}$ we let it stop; between $u=\frac{2}{3}$ and $u=1$ we let it resume its decelerated rate so that the third part of M_2^* takes place between $\frac{1}{4}$ and $\frac{1}{2}$. If the clocks are adjusted in this way, at every moment the positions of the particles are identical. Since we have achieved this identity by merely changing the time scales without ever reversing either clock we shall consider M_2 and M_2^* as geometrically equal.

In general, we shall say that two motions

$$M \text{ given by } X(t),\ Y(t) \text{ for } t_0\leq t\leq t_1,$$
$$M^* \text{ given by } X^*(u),\ Y^*(u) \text{ for } u_0\leq u\leq u_1,$$

are geometrically equal if, and only if, we can relate the moments between t_0 and t_1, and the moments between u_0 and u_1 in a nonrecurrent way and so that for every pair (t, u) of related moments we have

$$X(t)=X^*(u) \quad\text{and}\quad Y(t)=Y^*(u).$$

Here we say that the moments are related in a ***nonrecurrent*** way if

(1) every moment between t_0 and t_1 is related to at least one moment between u_0 and u_1; and every moment between u_0 and u_1 is related to at least one moment between t_1 and t_2;

(2) if t, u as well as t', u' are pairs of related moments, then the numbers $t-t'$ and $u-u'$ are not of opposite signs; that is, it never occurs that one of them is positive and the other one negative.

One readily derives from this definition that if t, u and t', u' are two pairs of related moments such that $u<u'$, then $t\leq t'$ and that from $t<t'$ it follows that $u\leq u'$. One further sees that t_0, u_0 as well as t_1, u_1 is a pair of related moments. If one moment of a rest-interval of either motion is related to a moment of the other motion, then every moment of the rest-interval is related to the same moment of the other motion. One can easily prove that for geometrically equal motions the paths traversed have the same length.

Now let Π and Π^* be two thermodynamical processes. For the sake of simplicity we shall assume that Π is determined by two functions $v(t)$ and $\theta(t)$ for $t_0\leq t\leq t_1$ and Π^* by two functions $v^*(t^*)$ and $\theta^*(t^*)$ for $t_0^*\leq t^*\leq t_1^*$. The extension of the subsequent remarks to processes whose characterization requires more than two functions presents no more difficulties than the extension of what has been said about motions in the plane to motions in a space of three or more dimensions.

We say that Π and Π^* are thermodynamically equal, if the motions in the (v, θ)-plane described by

$$v(t),\ \theta(t) \text{ for } t_0\leq t\leq t_1$$

and

$$v^*(t^*),\ \theta^*(t^*) \text{ for } t_0^*\leq t^*\leq t_1^*$$

are geometrically equal.

One readily proves that if the processes Π and Π^* are thermodynamically equal, then

$$\int_{t_0}^{t_1}P(v(t), \theta(t))dv(t)=\int_{t_0^*}^{t_1^*}P(v^*(t^*), \theta^*(t^*))dv^*(t^*).$$

Some scientists regard an interest in the history of their subject as mere antiquarianism, and it may be that the very remote past consists largely of mistakes to be avoided. But it deserves to be remembered that the history of any scientific discipline intimately determines the current modes of investigation. The frames of reference which appear eligible at any given epoch, the instruments accepted as respectable, and the types of "fact" taken to have evidential value are historically conditioned. To pretend otherwise is to claim for human reason, as manifested in scientific progress, a universality and fixity it has never manifested.—MAX BLACK, "The Definition of Scientific Method," *Science and Civilization*, Edited by Robert C. Stauffer (The University of Wisconsin Press, Madison, 1949).

Principles of Colorimetry

DONALD J. LOVELL
Naval Research Laboratory, Washington 20, D. C.

WE are all acutely conscious of the world of color about us, yet few of us can adequately describe it. If asked to define color, most physicists would likely resort to a description involving such quantities as wavelength and intensity; however, a chemist, a psychologist, or physiologist might describe color entirely differently. The chemist is conscious of color as a quality concerning a pigment or a dye. The psychologist and physiologist describe color in terms of visual perception.

Since very few textbooks treat color by a modern method,[1] the intent of this paper is to outline what the Committee on Colorimetry of the Optical Society of America and the International Commission on Illumination (ICI) have done, and what is being done, in presenting a unified description of color. Their task has been to harmonize the ways in which physicists and psychologists describe color. The first attempt to collect all of this information is being made by the Committee on Colorimetry of the OSA. Their report has been printed in part in the *Journal of the Optical Society of America* and the entire report is due to be published soon.

A Psychological Color Classification

The early studies of color classification brought forth what is known as a color sphere. Deane B. Judd, physicist for the National Bureau of Standards, describes the evolution of the color solid in about this manner: Assume that a man on a desert isle finds a trunk full of colored papers and attempts to classify the colors. After spreading them out in random fashion, he notices that several of them lack an important characteristic possessed by the others. They lack *hue*. Hue can be described by such familiar terms as red, orange, yellow, green, blue etc. The colors that lack hue are black, gray, and white.

As a result of this observation our hypothetical observer separates the colors into two classifications: the *chromatic colors* that have hue, and the *achromatic colors* that lack hue. He now turns to the chromatic colors and divides them into groups, putting all the reds together, all the blues together, and so on. He further notices that intermediate groups are found that form a continuous circle, ranging from red through orange to yellow, and then through green to blue and violet. Purple completes the circle back to red. This is classification by *hue*.

The achromatic colors are then separated into a single series from black through the grays to white. Now, examining all the colors of one hue, the observer notices that some of the colors are darker or lighter than others. Further, he finds that he can match this lightness to the grays of the achromatic colors, which he has just sorted. Thus, he finds the equivalent gray and classifies his colors by *brightness*. We commonly use this term to refer to a color as bright or dim.

When this is done, he notices another characteristic still to be considered. This characteristic is *saturation*. We think in terms of vivid, strong, or weak in this connection. For the achromatic colors the saturation is zero, while the saturation of the chromatic colors is greater than zero. If we consider a stimulus composed of both a chromatic and an achromatic color, the saturation is increased by increasing the amount of the chromatic color.

By these means our observer has divided the colors into various classifications. By placing all of the colors of the same brightness on one disk, with the hues placed consecutively around the disk, and the saturation increasing outward from the center, and similar disks of different degrees of brightness placed in order of that brightness above and below, the color solid is evolved. From this experiment, we learn the three attributes of color that will serve to identify it by color sensation. These are *hue*, *brightness*, and *saturation*.

Many such color solids have been assembled for matching a sample by visual comparison. Familiar solids have been prepared by Munsell, Maerz and Paul, the Bureau of Standards,

[1] Notable exceptions are found in the last two references in the bibliography at the end of this paper.

Ostwald, and many others. All of these systems depend upon obtaining a visual judgment made with the unaided eye of hue, of brightness, and of saturation. Experience shows that there is a divergence in the color judgments obtained from several observers. Thus, when accuracy is desired, the opinions of a multitude of observers must be averaged. To avoid this, methods of specifying colors have been devised whose results are expressed in terms of the judgment which would be made by a standard observer. In these spectrophotometric methods, visual observations are eliminated or reduced to chromatic brightness matchings. Spectrophotometry gives a physical basis for the definition of color in terms of the judgment of a "standard observer."

The Young-Helmholtz Theory

It is a well-known fact that the eye is not a selective instrument. The ear can detect two separate vibrations as such. That is, if a violin and a trumpet are sounded simultaneously, the ear has little difficulty in recognizing that two instruments are emitting sound. On the other hand, the eye cannot distinguish two superimposed colors as such. Taking advantage of this fact Young, and later Helmholtz, found that it is possible to color-match any given light, using a combination of three primary sources.

What Young and Helmholtz did, then, was to show that with three primaries, any given color could be matched. It was learned that occasionally one color was found that could not be matched by direct addition of the three primaries. However, it was always found that if one of the primaries was added to the given color, then the other two primaries would produce a color match with the combination of the sample and the third primary.

Consider that this experiment is repeated using three monochromatic primaries of, say 450 mμ, 550 mμ, and 620 mμ. We will illuminate one side of a photometer field with monochromatic light, varying it through the visible spectrum. Throughout this experiment the energy of the monochromatic source is to be held constant while its wavelength is varied.

We begin the experiment by choosing the monochromatic source to be of the same wavelength as one of the primaries, say 620 mμ. We will find that a certain amount of that one primary and zero amounts of the other two are then needed for a color match. We will set the value of the primary used at 100 for convenient comparison purposes. This procedure will then be repeated at the two other primary wavelengths, again choosing the amount of the required primary at 100, and the other two zero, of course.

Following this, the monochromatic source is varied from wavelength to wavelength throughout the visible spectrum, and the values of the three primaries required to produce a color match are determined. Figure 1 shows the results of this experiment as they would be obtained by a standard observer. The amounts of the three primaries required to reproduce a spectrum color are represented by the letters α, β, and γ; and are called *tristimulus values*.

It is to be noted that each of these curves occasionally dips below the horizontal axis, or has negative values. These are the regions where that primary must be added to the monochromatic source, and the color match is then effected by the other two primaries.

The interpretation of color vision by the Young-Helmholtz theory is that the eye contains

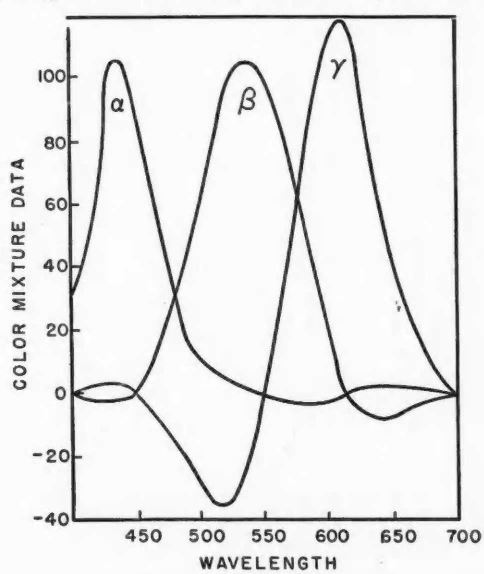

Fig. 1. Color mixture data for monochromatic primaries of wavelength 450 mμ, 550 mμ, and 620 mμ.

three independent color-sensitive receptors. On this basis, the three curves of Fig. 1 might be regarded as representing the individual spectral sensitivities of the three radiation detectors in the eye. The negative portions of the curve would then be assumed to be places where the sensation is inhibited rather than stimulated by that wavelength. Maxwell carried out experiments in the attempt to verify the Young-Helmholtz theory in 1854. He was followed by Koenig and Dieterici in 1892 and by Abney in 1913. This work, however, was of only academic interest until 1922. At that time, the colorimetry committee of the OSA summarized and republished these results. This enabled the physicist to base tristimulus specifications on spectrophotometric data, and to avoid many of the uncertainties of colorimetry.

Evaluation of the Tristimulus Values

In 1928, both Wright and Guild, working independently in England, redetermined these fundamental data, using carefully picked observers. They used different primaries, but when the work was reduced to a comparable basis the two methods agreed remarkably well.

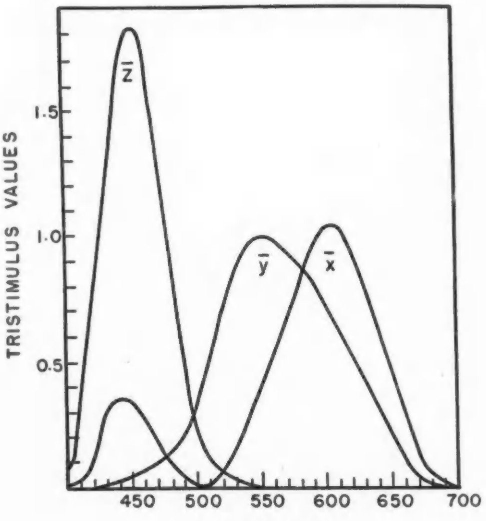

FIG. 2. Tristimulus values for the various spectrum colors. The values $\bar{x}$, $\bar{y}$, $\bar{z}$ are the amounts of the three ICI primaries required to color-match a unit amount of energy having the indicated wavelength.

There appears to be no anatomical evidence to verify the Young-Helmholtz theory. Further, we have selected our three primaries quite arbitrarily. Notwithstanding, this theory of three independent color-sensitive receptors gives exceptionally fine results.

The selection of the three primaries is clearly quite arbitrary. However, it can be shown that a linear relationship exists between the respective tristimulus values of the primaries used by different observers. It would be convenient to use primaries which yield tristimulus values which are positive throughout. However, no such curves are to be found experimentally. Nevertheless, applying a mathematical transformation that exists between two sets of primaries,

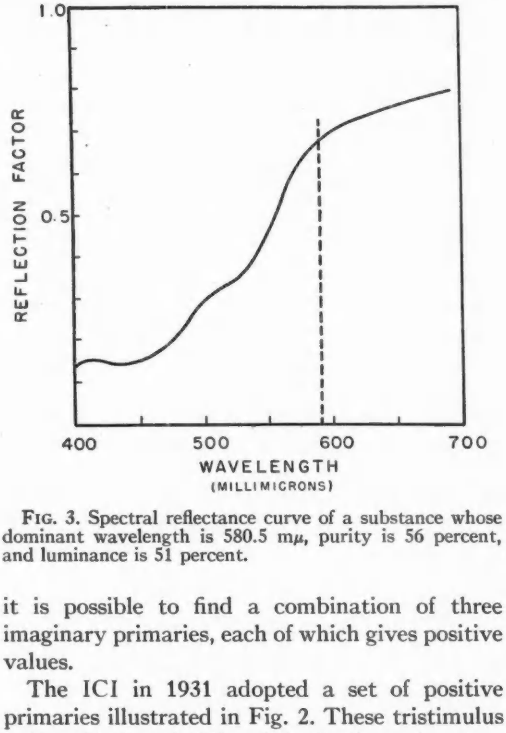

FIG. 3. Spectral reflectance curve of a substance whose dominant wavelength is 580.5 mμ, purity is 56 percent, and luminance is 51 percent.

it is possible to find a combination of three imaginary primaries, each of which gives positive values.

The ICI in 1931 adopted a set of positive primaries illustrated in Fig. 2. These tristimulus values ($\bar{x}$, $\bar{y}$, and $\bar{z}$) have the further advantage that the $\bar{y}$ values correspond to the luminosity curve of the human eye. This advantage will soon be discussed.

A psychological, or subjective, interpretation of these three tristimulus primaries can be given in the following manner. The value of $\bar{x}$ repre-

sents a primary which is a reddish purple of higher saturation than any obtainable color of this hue. The value of $\bar{y}$ represents a green primary whose dominant wavelength is 520-mμ, but of considerably more saturation than any spectrum color. The value of $\bar{z}$ represents a blue primary that is considerably more saturated than the spectrum color whose wavelength is 477 mμ.

If the sample stimulus is not monochromatic (which is the usual case), these curves can still

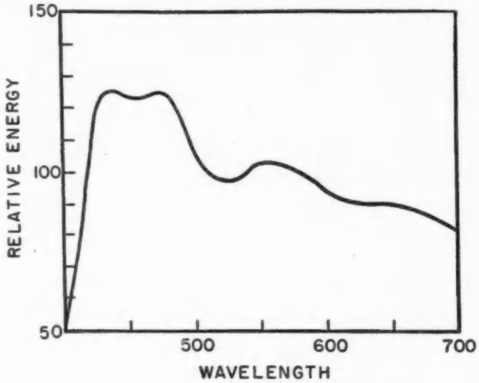

FIG. 4. Relative spectral distribution of the energy radiated per unit time by ICI illuminant "C"

be used to form the desired computations. Calling the energy distribution of the sample stimulus E_λ, the amount of each primary required for a color match ($\bar{x}$, $\bar{y}$, or $\bar{z}$, respectively) can be found by multiplying E_λ by the corresponding values of $\bar{x}$, $\bar{y}$, and $\bar{z}$ for each wavelength, and summing the results. We can, therefore, represent this mathematically by the following integrals:

$$X=\int_0^\infty \bar{x}E_\lambda d\lambda$$

$$Y=\int_0^\infty \bar{y}E_\lambda d\lambda$$

$$Z=\int_0^\infty \bar{z}E_\lambda d\lambda.$$

Frequently, the stimulus E_λ represents radiation reflected or transmitted by a material, the color of which we are interested in measuring. Both transmitted and reflected light can be

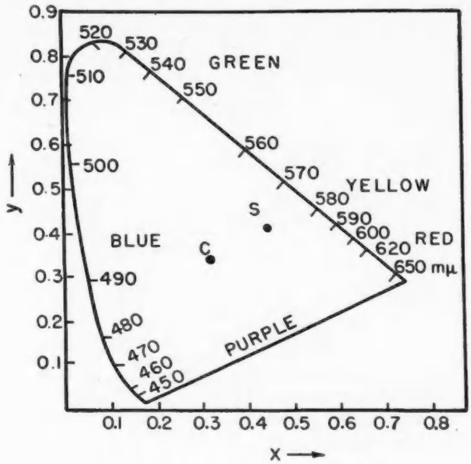

FIG. 5. Chromaticity diagram.

treated in the same way; so for convenience, we will confine our discussion to reflected light.

We may determine the spectral reflectance of a given sample directly with the use of a spectrophotometer. Such a response is given by Fig. 3 for a particular sample. The color will still depend on the choice of the light source so the ICI has adopted standards of illumination. The one most generally used is ICI illuminant "C" which approximates average daylight. The Bureau of Standards provides this source in the form of a tungsten lamp operating at a color temperature of 2838°K and provided with a suitable filter. The relative spectral distribution of such a source is given by Fig. 4.

If the sample of Fig. 3 then, is illuminated by illuminant "C," we may represent the resultant reflected energy as the sum of the products of R (the reflectance of the sample) and $E_c(\lambda)$ (ICI illuminant "C") taken wavelength by wavelength. Thus, at any wavelength, the reflected energy is that of illuminant "C" as modified by the reflectance of the sample. We may replace E_λ in the above equation by $RE_c(\lambda)$, giving us:

$$X=\int_0^\infty \bar{x}RE_c(\lambda)d\lambda$$

$$Y=\int_0^\infty \bar{y}RE_c(\lambda)d\lambda$$

$$Z=\int_0^\infty \bar{z}RE_c(\lambda)d\lambda.$$

It is obvious that the exact evaluation of the above integrals will present a difficult problem, so we approximate the desired results by taking finite sums. The above equations then become:

$$X=\sum_{380}^{780\ \mathrm{m}\mu}\bar{x}RE_c(\lambda)\Delta\lambda$$

$$Y=\sum_{380}^{780}\bar{y}RE_c(\lambda)\Delta\lambda$$

$$Z=\sum_{380}^{780}\bar{z}RE_c(\lambda)\Delta\lambda.$$

Chromaticity Coordinates

Applying the rarely used method of weighted ordinates, the intervals of $\Delta\lambda$ are made equal in computing the sums. The values of $E_c\bar{x}$, $E_c\bar{y}$, and $E_c\bar{z}$ are to be found tabulated in the *Handbook of Colorimetry*.[2] The values of R at the midpoint of each $\Delta\lambda$ group are then chosen from the reflectance curve and multiplied by the proper $E_c\bar{x}$ factor. Summing these provides a figure proportional to X. Similarly, the values of Y and Z may be found.

More frequently the method of selected ordinates is used. The products, $E_c(\lambda)\bar{x}\Delta\lambda$ are summed up and the total divided into a number of equal divisions. By this method the wavelength scale is broken down into sections containing equal fractions of the total weighted energy. For the wavelengths of the midpoints of each section, corresponding values on the reflectance curve can then be found. The sum of these values is proportional to X and the proportionality factor is found in the *Handbook of Colorimetry*. Similarly, the values of Y and Z are found by this method. The advantage of this system is, of course, that fewer computations are needed. The wavelengths selected for this method are also tabulated in the *Handbook of Colorimetry*.

Having found the tristimulus values X, Y, and Z, we are now prepared to define and evaluate the *chromaticity coordinates* (or trichromatic coefficients, as they were called). These are defined as the fraction of the three color standards required to produce a color match. These can be expressed as:

$$x=X/(X+Y+Z)$$
$$y=Y/(X+Y+Z)$$
$$z=Z/(X+Y+Z).$$

Since $x+y+z=1$, only two of these quantities are required to determine specifically a color match. Furthermore, by choosing two of these variables, say x and y, as Cartesian coordinates, we may locate a color graphically.

A Spectrophotometric Description of Color

If we were to plot the chromaticity coordinates (x and y) of the spectral colors on such Cartesian coordinates, we would obtain the horseshoe-shaped diagram shown in Fig. 5. This is called a *chromaticity diagram*. All real colors must lie within this boundary. Further, it can be shown that any combination of two colors lies on a straight line joining these colors on the chromaticity diagram.

We may also plot the chromatic coordinates of the ICI illuminants on this diagram. The ICI illuminant "*C*" is designated on the chromaticity diagram of Fig. 5 by the letter *C*.

The chromatic coordinates of the sample whose reflectance curve is illustrated in Fig. 3 are located by the letter *S* on this diagram. This point *S* may be considered as a mixture of the ICI illuminant "*C*" and a pure color. The spectral color thus designated is referred to as the *dominant wavelength* of that particular color. In the case of the sample shown, the dominant wavelength is 580 mμ. Dominant wavelength, as defined here, is seen to be comparable to hue as discussed earlier in this report.

This applies to any of the colors except the purples. The purples may be designated by their complementary color, or complementary dominant wavelength. Since complementary colors are defined as those which when added in the proper proportions will produce white, we find the complementary wavelength by the continuation of the line joining the white illuminant and the sample to the other side of the spectrum locus. Complementary colors thus will lie on a straight line through the point of illuminant "*C*."

A second characteristic of color that evolves from the chromaticity diagram is *purity*. This characteristic is defined as the ratio of the length

[2] A. C. Hardy, *Handbook of colorimetry* (Technology Press, Cambridge, Massachusetts, 1936).

of line connecting the illuminant point and the sample point to the length of line connecting the illuminant point and the spectrum locus when drawn through the sample point. We can express this by the equation:

$$P=(Y_s-Y_c)/(Y_l-Y_c)=(X_s-X_c)/(X_l-X_c),$$

where the subscripts s stand for sample, c for illuminant "C" and l for the spectrum locus; and the x's and y's are the chromatic coordinates designated by these subscripts. Purity as now defined is seen to be comparable with saturation as discussed previously. We see that a pure spectral color has a purity of 100 percent, all other colors less, while the grays, black, and white have zero purity.

We have not yet completely determined a color match by these two dimensions. It has been shown that it was also necessary to classify colors by their brightness. Since the tristimulus values $\bar{y}$ was chosen to coincide with the luminosity curve of the human eye, we may compare the value of Y appropriate to the reflectance shown in Fig. 3 to corresponding values of Y that would be obtained if the sample were a perfect reflector. The quotient thus formed is a measure of the *luminous reflectance* of the substance *measured*, and is comparable to the brightness as discussed earlier.

From these considerations, we find the sample of Fig. 3 to have the following characteristics:

Dominant wavelength	580 mμ
Purity	56 percent
Luminous reflectance	51 percent.

We see now that it is possible to measure a color in terms of these three characteristics. They are determined from the tristimulus values, which in turn are determined from the reflectance or transmittance curve of the sample. Spectrophotometers have been built to draw curves automatically, and to evaluate the tristimulus values while in operation. The chromatic coordinates can easily be computed from these tristimulus values to determine the dominant wavelength and purity by means of the charts in the *Handbook of Colorimetry*.

The value of Y is already proportional to the luminous reflectance. Thus, the characteristics of color (dominant wavelength, purity, and luminous reflectance) can be determined quickly from the curve given by a recording spectrophotometer.

Summary

In conclusion, let us review what color means in terms of this paper. Color is an index to that capacity of an object which modifies the visual sensation produced by incident light. This modification depends upon the selective absorption of the object. Color therefore depends upon the spectral distribution of the incident radiant energy, the selective absorption of the reflector, and upon the psychophysical functions of human vision.

The Committee on Colorimetry of the Optical Society of America defined light as "that aspect of radiant energy of which an observer is aware through the visual sensations which arise from the stimulation of the retina of the human eye." This is a psychophysical concept. Upon this basis, "Color consists of the characteristics of light other than spatial and temporal inhomogeneities."

The characteristics of color are determined in terms of: (1) the dominant wavelength, (2) appropriate photometric quality, and (3) purity.

Bibliography

American Society for Testing Materials, *Symposium on Color*, 1941.

Report of the Committee on Colorimetry, *J. Opt. Soc. Am.* **33**, 544–554 (1943); **34**, 183–218, 245–226, 633–688 (1944); **35**, 1–25 (1945).

Evans, Ralph M., *An Introduction to Color* (John Wiley & Sons, New York, 1948).

Hardy, A. C., *Handbook of Colorimetry* (Technology Press, Cambridge, Massachusetts, 1936).

Hardy, A. C., "The history and design of the recording spectrophotometer," *J. Opt. Soc. Am.* **28**, 360 (1938).

MacAdam, David L., "Color vision," *Physics Today* **1**, 10 (May, 1948).

Murray and Spencer, *Colour in Theory and Practice* (Chapman and Hall, London, 1939).

Sears, Francis W., *Principles of Physics* (Addison-Wesley, Cambridge, Massachusetts, 1946), Vol. III.

The Vector Cross Product in Elementary Electrodynamics

M. E. GARDNER
University of California, Davis, California

MOST textbooks of college physics do not use the qualitative concept of the vector cross product for describing the interactions of electric currents with magnetic fields. The introduction of the concept is feasible and desirable. The consistent use of the concept simplifies the teaching of these phenomena and reduces the number of rules a student is required to remember. The easily confused right-hand and left-hand rules for motors and generators can be replaced by a simpler and more logical rule. The concept can be developed in successive stages, beginning with the right-hand screw rule for the determination of the magnetic field about a wire conducting current.

Increased emphasis is being placed upon the fact that an electric current in a metallic conductor consists of a flow of electrons from points of low potential to points of higher potential. However, it may be pointed out to the student that for the purpose of analysis it is immaterial whether a current is considered as a flow of electrons or whether it is considered as a flow of positive charges from points of high potential to points of lower potential. By retaining the classical and conventional description of an electric current, the concept of the vector cross product can be profitably employed. The method of developing the concept is outlined here, with accompanying illustrations.

Electromagnetic phenomena are first encountered with the fact that magnetic fields are associated with electric currents. When electrical charges flow along a wire the direction of the magnetic flux lines around the wire due to the current can be determined by the right-hand screw rule, as is given in many textbooks. If the screw shown in Fig. 1 advances in the direction of the positive current, then the direction of rotation of the screw gives the direction of the magnetic lines of force around the wire. If now the wire is bent into the form of a loop, it is seen that by the application of the right-hand screw rule, all the magnetic lines of force enter at one face and emerge from the other. A simple extension of the right-hand screw rule shows a reciprocal nature; by rotating the screw in the direction of the current, the screw advances in the direction of the magnetic lines of force. This latter form of the rule is particularly useful for finding the magnetic poles of a solenoid conducting current. Since conductors in the form of coils of wire or solenoids become magnets when conducting currents, then when they are placed in magnetic fields and are free to rotate, such as in d'Arsonval galvanometers, the direction of rotation of the coils becomes apparent.

The usual description of the interaction of electric currents with magnetic fields employs the Faraday concept of lines of magnetic field

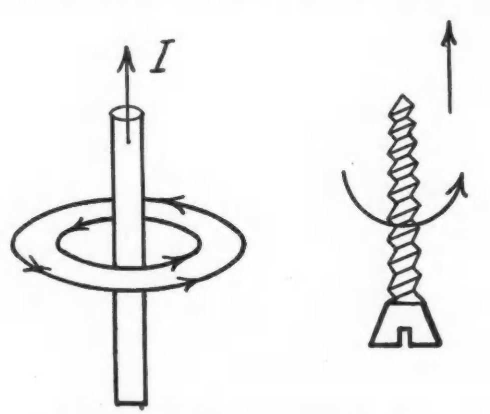

FIG. 1. Right-hand screw rule applied to a conductor carrying current.

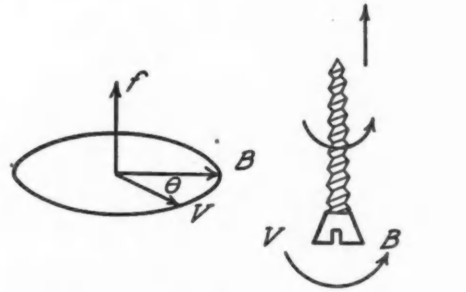

FIG. 2. Right-hand screw rule applied to electrical charge moving in a magnetic field.

strength acting like stretched rubber bands. This concept is useful, but by a slight extension of the familiar right-hand screw rule, the concept of the vector cross product may be introduced which gives the interactions of electric currents with magnetic fields directly. This interaction may be considered from the point of view of the individual free charges in a conductor.

If there are n free charges per unit volume in a wire, each of charge q, traveling with a speed v along the wire of cross-sectional area A, then the current will be $I=nqvA$. Each charge in motion contributes to the magnetic field whose direction is given by the right-hand screw rule. If the moving electrical charge is in a magnetic field, there will be an interaction between this magnetic field and the field of the moving charge.

If the flux density of the field is B, the force exerted on the moving electrical charge will be $f=qvB\sin\theta/10$, where θ is the angle between v and B. The force f is in dynes when q is in abcoulombs, B is in gauss, and v is in centimeters per second. The direction of the force is given by the right-hand screw rule. If the direction of v is turned into B through the smaller angle of Fig. 2, then f is the direction a right-hand screw would advance. Since the force f is perpendicular to the plane determined by v and B, it is a centripetal force. When $\theta=90°$ the path of the charged particle is a circle whose radius is given by $qvB=mv^2/R$. The cyclotron is an application of this condition.

When a wire conducting a current I is in a magnetic field of flux density B, each moving charge q is subjected to the force f. The number of charges in a length L of wire is $N=nLA$. The total force on the charges is

$$F=Nf=\frac{nLAqvB\sin\theta}{10}$$

or

$$F=\frac{ILB\sin\theta}{10}. \qquad (1)$$

The direction of F is given by the right-hand screw rule (vector cross product) and F is perpendicular to the plane determined by I and B as shown in Fig. 3.

When a loop of wire conducting a current is in a magnetic field, the right-hand screw rule can

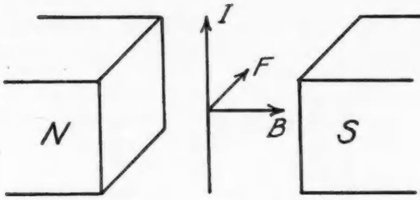

FIG. 3. Force exerted on a conductor in a magnetic field.

be used in either of two ways to obtain the motion of the coil. The direction of F on the sides of the coil can be obtained as shown in Fig. 3, or the screw can be rotated in the direction of the current to obtain the direction of the flux of lines emerging from the face of the coil.

If a straight wire conducting a current I_2 is in the vicinity of another straight wire conducting a current I_1, the direction of the force on the conductors can be obtained by application of the right-hand screw rules. The direction of B_1 is obtained from the first form of the rule, and the magnitude of B_1 is given by the law of Biot and Savart. The direction of the force F on the current I_2 is given by the cross product, and for currents in the same direction as illustrated in Fig. 4, the force is one of attraction. The magnitude of the force is given by Eq. (1).

If a loop of wire is rotated in a magnetic field by some mechanical means so that the sides move in the direction indicated by v in Fig. 5, then every charge q in the wire experiences a force $f=qvB\sin\theta$. The direction of f is perpendicular to the plane containing v and B and, by the right-hand screw rule, is the direction indicated by i_g in Fig. 5. Consequently, if the ends of the wire are connected by a resistor the charges q

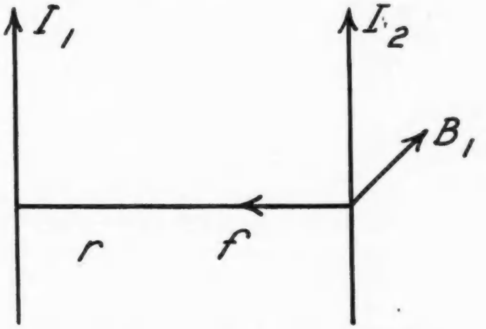

FIG. 4. Interaction force between parallel conductors.

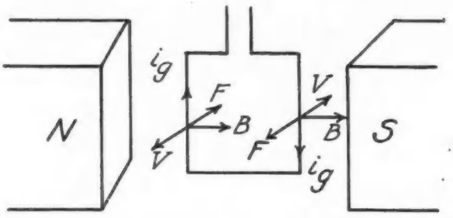

FIG. 5. Right-hand screw rule applied to a coil carrying a current in a magnetic field.

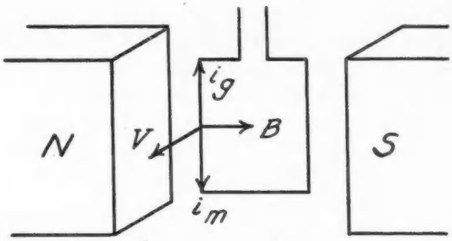

FIG. 6. Torque exerted on a coil carrying a current in a magnetic field.

will move with a speed u along the wire. But this constitutes a current $i_g = nquA$ along the wire. A wire conducting a current i_g in a magnetic field is subjected to a force F whose direction, as given by the vector cross product, is 180° from v. Thus the mechanical force F' required to maintain the speed v of the wire must be equal and opposite to F. The existence of the force F is convincingly demonstrated by using a hand cranked generator or magneto, and alternately, opening and shorting the output.

When a current i_m is sent through a coil as shown in Fig. 6 and the coil is free to rotate in a magnetic field, the moving charges represented by the current i_m will have a force exerted on them. The direction of this force, as given by the vector cross product, is the direction v. Hence a torque is exerted on the coil causing it to rotate. If the arrangement is such that the coil can be made to rotate continuously, a motor is obtained. But if electrical charges move through a magnetic field in the direction v, a force will be exerted on them. The direction of this force as given by the right-hand screw rule is designated by i_g. The emf generated by the coil rotating in the magnetic field tends to produce a generated current i_g in opposition to the motor current i_m so that the net current into the motor may be regarded as $i = i_m - i_g$. This is the back emf generated by the motor.

When a bar magnet is brought up to a coil of wire, the direction of the current can be readily determined, for if a right-hand screw is advanced in the direction of the magnetic lines of force in the coil, the rotation of the screw gives the direction of the current in the coil windings. The vector cross product can also be used for explaining the skin effect resulting from the self-inductance of a wire. As the magnetic field is being established and moves outward from the wire, this motion is equivalent to having electrical charges move in the opposite direction. The vector cross product shows that a back emf must be generated.

The right-hand screw rule may be summarized as follows:

(1) When the screw advances in the direction of the current, the rotation of the screw gives the direction of the magnetic lines of force.

(2) When the screw is rotated in the direction of the current in a coil, the screw advances in the direction of the magnetic lines of force.

(3) The direction of the force on positive charges moving in a magnetic field is given by the advance of a screw when the screw is rotated such that the direction of motion of the charges is turned towards the direction of the magnetic flux through the smaller angle, i.e., v towards B for θ less than 180°.

In other words, we must set our sights not on what we would like to know, but first on what we do know with certainty.—MAX PLANCK, The Meaning and Limits of Exact Science. (*Science* **110**, 319 (Sept. 30, 1949)).

LETTERS TO THE EDITOR

Competitive Test for High School Students in Southern California

ONE objective of the Southern California Section of the AAPT has been to encourage a closer relationship and understanding between the teachers of physics in high schools and colleges. Soon after the formation of the section it was agreed that an annual competitive physics test for high school students sponsored by the section would do much to promote an interest in physics on the part of students and would at the same time furnish a definite tie between the section and the high school teachers in the area. It was believed that the content of the examination would indicate to the high schools the scope of the material which should normally be covered in a thorough high school course in physics. In turn it was felt that the comment and discussion resulting from the analysis of the examination content would yield considerable information to the college group of the problems existing in high school teaching. There was never any intention on the part of the section to intimidate or embarrass any high school or any high school instructor by showing up the lack of preparation of students from a given origin. Rather it was intended that the examination would furnish a medium of comparison of accomplishments on the part of all schools in the Southern California area and the means whereby constructive criticism could be presented and improvement secured.

With the above objectives in mind the first annual high school test was held on June 2, 1945, less than six months after the organization of the section. The participation of 91 students from 29 different high schools gave great encouragement to the project. This response along with the endorsement of many high school teachers launched the venture as an annual event. By action of the Executive Committee of the section, suitable certificates indicating the nature of the award and competition were secured and presented to the first ten students in the order of their standing in the examination.

With some experience and considerably more time for organization the second annual test was held on June 1, 1946. Scholarships covering a full year's tuition were secured from Pomona College, Occidental College, Redlands University, and the University of California at Los Angeles. The total value of these awards was over $1500 to be assigned on the basis of standing in the examination. Additional annual scholarships are now available from the California Institute of Technology, University of Southern California, and Whittier College in addition to those first offered in 1946. These scholarships are available through the courtesy of the colleges and universities mentioned and have a total annual value of over $2500. The assignment of the scholarships is controlled by the Executive Committee of the section on the basis of the standings as reported by the Test Committee. Participants must indicate their choice of scholarship on the first page of the examination. In general the policy has been to assign scholarships only to those students ranking in the first fifteen, who receive certificates of award from the section. Some idea of the extent of the participation may be gained from Table I: In order to encourage wide participation throughout the Southern California region the examination is given at Redlands, Santa Barbara, and San Diego at the same time as the main group is taking the test in the Los Angeles area. All high school instructors are invited to attend a meeting arranged especially for them at the host institution in the Los Angeles area during the progress of the test. Copies of the test are distributed and discussed. Suggestions and criticisms are noted for use in preparation of the next year's examination.

The form of the examination has varied somewhat from year to year. The length of the examination has been limited to two hours and has at times included matching test on definitions, true-false questions, multiple-choice questions, and problems covering the general field of physics. Main emphasis has always been on the solution of problems which are graded on the basis of method as well as on accuracy of the answers. The 1949 test consisted of sixteen selected problems in the fields of mechanics of solids, liquids, and gases, in sound, in heat, in electricity, and in light. Students are permitted the use of slide rules. All necessary physical constants (and some superfluous ones) are tabulated on the first page of the examination. The tests are mimeographed and adequate space is provided below the statement of each problem for the working of each problem.

In 1947 the test participation was on a team basis similar to the test sponsored annually by the American Chemical Society. The Test Committee has abandoned this procedure as undesirable. It is felt that wider participation occurs when the competition is open to all students who wish to enter. It is also believed that concentration on a few students by the high school teacher is less likely to occur. Through this means of wider competition it is hoped that many more high school students are stimulated to higher proficiency in physics. It is also believed to be a more democratic approach.

As the range of participation in the Test became extended and its reputation established it became evident to the Executive Committee of the section that a "permanent" Test Committee should be established in order to insure continuity of procedure. Such a committee was established in 1948 and consists of five members elected by the section who serve for terms of three years. The

TABLE I.

Year	Schools represented	Students participating
1945	29	91
1946	31	124
1947	43	138
1948	47	193
1949	51	182

members of the Test Committee are from college and university faculties. High school teachers are not eligible. The Test Committee is responsible for the release of announcements concerning the nature and time of the examination, for the composition of the examination, and for its grading. The Committee reports the results of the examination to the Executive Committee of the section which assigns all scholarships and sends the certificates of award to the top fifteen contestants along with copies to their respective schools.

The experience of the last five years indicates clearly that the sponsorship of the annual competitive physics test for high school students by the section is of great value. The original objectives which gave incentive for establishment of this activity have largely been realized. The carrying forward of the project from year to year represents considerable work on the part of the Test Committee. The effort seems justified on the basis of the results being achieved. It is felt that the test has brought the field of physics to the attention of many high school students and that it has in special cases been of great assistance to worthy students through the generous scholarship grants made available through the competition. Although no formal study has been made of the college records of scholarship recipients, an informal study has indicated that they have in general done very well. It is probably safe to say that the section is making a great contribution to physics in the Southern California area by the sponsorship of the annual high school physics test.

V. L. BOLLMAN

Occidental College
Los Angeles, California

The Pronunciation of *Electricity*

IN his discussion of the pronunciation of *electricity*,[1] Professor Perkins reveals a regrettable misconception. Webster's *New International Dictionary* lists the following accepted pronunciations of *electricity*, *electrician*, and *electrical:*

e-lek″tris′i-ti or el″ek-tris′i-ti
e-lek″trish′an or el″ek-trish′an
e-lek′tri-kal.

The sound of the initial *e* ("long" in the entries on the left, "short" in those on the right) is dependent on the position of the secondary accent (″), and recession of this accent to the first syllable is in accordance with the general tendency in English; see Webster, Section 63. In the third word this possibility does not exist, because no syllable receives a secondary accent. Therefore, what Professor Perkins was told at Columbia is correct, but it implies no disparagement either of the man who repairs the doorbell or of the man who prefers to say el″ek-tris′i-ti.

"Long" and "short" applied to classical vowels or syllables refer to *quantities* (durations in time), not to *qualities* of sound as in English, and there is no direct correlation between the classical quantities and the English qualities.[2] *Library* comes from *liber* (*i* short), "book," and *liberty* comes from *liber* (*i* long), "free." Other examples of the lack of correlation will be found in Fowler's *Dictionary of Modern English Usage* under False quantity. The quantities are relevant to English pronunciation only so far as they determine the accents, which in turn partly determine the English vowel qualities. In *electricity* the accents are determined by other factors than the quantity of the initial Greek vowel; knowledge that this was *eta* (long e) is therefore of no value in deciding the "correct" English pronunciation.

WILLIAM FULLER BROWN, JR.

Sun Physical Laboratory
Newtown Square, Pennsylvania

[1] Henry A. Perkins, *Am. J. Physics* **17**, 398 (1949).
[2] The rules for pronunciation of classical words in English are summarized in Webster's *New International Dictionary* (ed. 2), Sections 269–270. More complete rules are given in old Latin grammars, such as that of Harkness (1881). These rules apply strictly only to words taken over bodily into English; when the ending changes, as in *electricity*, other changes may occur, and in any case the ultimate criterion is usage.

Larmor's Theorem in Quantum Mechanics

IT will be shown below that Larmor's theorem may be used in quantum mechanics and gives the correct formula for the splitting of energy levels in the normal Zeeman effect.

Consider a particle of charge e and rest-mass m acted upon by a uniform weak magnetic field H directed along the z axis. In classical electrodynamics Larmor's theorem states that the motion of the particle is approximately the same as if the magnetic field were removed but the x and y axes rotated about the z axis with angular velocity $eH/2mc$. Assume now that H can be introduced classically into quantum mechanics. Then its replacement by the Larmor rotation means that the eigenstates of x and y change their significance, so that the corresponding basic vectors $|xy)$ in function space change. We may therefore write

$$|xy)=U^{-1}|x_0y_0), \tag{1}$$

where U is a certain time-dependent unitary operator and $|x_0y_0)$ is equal to $|xy)$ at the instant the field is switched on. Function U may be expressed as

$$U=e^{-iAt/\hbar}, \tag{2}$$

where A is a certain Hermitian operator. In order to correspond to the periodicity of the Larmor rotation, U must be a periodic function of t, whence A must be time-independent, and the period must be $2mc/neH$ where n is an integer.

In the usual quantum theory of the Zeeman effect,[1] H is used classically in the Hamiltonian of the atomic system. We are therefore justified in using Larmor's theorem as described above. In the absence of an external field, the normalized state vector of the system $|p_t)$ is related to its value $|p_0)$ at zero time by

$$|p_t)=e^{-iEt/\hbar}|p_0), \tag{3}$$

where E is the Hamiltonian operator and is time-independent. Upon switching on the magnetic field the basis $|xy)$ starts to move according to Eq. (1). This motion may, however, be changed to that of an extra motion of $|P_t)$

by means of a contact transformation, in a manner analogous to the transition from the Heisenberg picture to the Schrödinger picture. We then have

$$|p_t') = U|p_t) = e^{-iAt/\hbar}e^{-iEt/\hbar}|p_0). \quad (4)$$

Suppose that the operation U is applied first, then one may obtain $|p_t')$ from $|p_0)$ by

$$|p_t') = e^{-iFt/\hbar}e^{-iAt/\hbar}|p_0), \quad (5)$$

where F is a certain Hermitian operator, and by equating the time derivatives of the right-hand sides of Eqs. (4) and (5), we find that $E=F$, so that A and E commute. Now if E' is the Hamiltonian in the presence of the field, it is constant since the field is constant, and

$$|p_t') = e^{-iE't/\hbar}|p_0), \quad (6)$$

whence it follows from Eqs. (4) and (6) that

$$E' = E + A. \quad (7)$$

Since E and A commute, they can be simultaneously diagonalized, and using the periodicity of U discussed above, the eigenvalues of E' corresponding to the mth eigenvalue of E are found to be

$$E_{m'}' = E_m + neH\hbar/2mc, \quad (8)$$

which is the correct formula for the energy levels in the normal Zeeman effect.

This method has the advantage that one does not require explicitly the theory of angular momentum operators or of the Hamiltonian including an electromagnetic field, but has the disadvantage that we cannot carry it further to give the frequencies and polarizations of spectrum lines since our limited knowledge of A prevents us from finding the appropriate selection rules.

R. CADE

University College of the West Indies, Jamaica, B. W. I.

[1] P. A. M. Dirac, *Quantum mechanics* (Oxford University Press, New York, 1947), p. 165.

A Device for Showing Vectors in Space

IN the first course in analytical mechanics students invariably experience difficulty in visualizing vectors in space. Even a skillful drawing in colored chalk often fails to show the student who is lacking in spatial perception how a vector lies, say, in the seventh octant. To assist the instruction of vectors in three dimensions the writer has built himself a "toy" which has shown itself to be highly effective.

A cubical wooden block about 3 inches on an edge is drilled with small holes as follows: (1) Holes through the centers of the plane faces; (2) holes through the midpoints of the edges; (3) holes through the corners. All these holes are drilled so as to pass through the geometric center of the block.

In drilling extreme care must be exercised to insure that the holes pass through the center of the block. In this connection, the writer found the job a troublesome one for shop students, since the diagonal of a cube does not make 45° with the diagonal of a face drawn from the same vertex.

The face holes are now fitted with coordinate axes—long sticks appropriately painted, and the other holes are fitted with shorter "vectors," pointed and painted. A larger rigid negative y axis provides the upright support, the lower end of which may be imbedded in a heavy base. With all the vectors in place the figure is attractive and instructional. The octants are immediately demonstrable and the spatial configuration is much clearer to the student.

JULIUS SUMNER MILLER

Dillard University
New Orleans, Louisiana

Concerning Historical References in General Physics

AN examination of the general physics texts published over the years shows some authors given to historical references and elaborations thereon, and others who, to put it in the vernacular, stick strictly to business. Some, indeed, indicate in their prefaces that they have done a bit with the history of the subject; others definitively state that they will have none of it. I believe that a goodly amount of historical material, source literature, and collateral reading should be put into a first course in physics, whether it be a terminal course for arts students or a technical course for physicists and engineers. Some of the historical and humanistic background of the subject sometimes gets into courses for nonscience majors but those who will teach and work in the sciences know pitifully little about its history and the men who made it. It is claimed, of course, that historical material and humanistic references make a course "easier," but these things can be kept on the same intellectual plane as the purely analytical material. Truth to tell, the subject needs a little leaven now and then, even for the technical man, and historical references help supply it. Furthermore, the subject of physics might well be presented as the contributions and achievements of men, which is what it is, indeed. And this approach is inspiring if properly presented. Above all, the discussion of historical periods and personalities gives the student a larger over-all point of view, which, most teachers must admit, few physics students ever acquire in their student days. I have had excellent students (in subject-matter proper) who could hardly name a man or event in the decade 1895 to 1905, and these ten years are monumental. Physics students ought to come away from a course knowing something about J. J. Thomson, Becquerel, Roentgen, Planck. They ought, too, to know something about Bohr and Pauli and Fermi and Dirac. But a recitation of names only is not what I mean. I have had contact recently with a number of good graduate students in physics, few of whom knew anything of the history of their subject. Many could not place Maxwell in his appropriate place in the development of physical thought, to take an example at random.

The argument against historical material and source literature and quotations is simple enough to state. There is no time, the teachers say. But I dare suggest that a measure of values is here involved, and this position is, I

believe, quite unassailable. Students should at least know some men by their contributions and some contributions by the men who formulated them. That is, students should be able to associate men and ideas and place them with reasonable accuracy in their proper historical place. Toward this end I pursue a few elementary schemes in my general physics course which pays off large dividends. Whenever a name or fact arises which deserves historical comment, I make it. I give references and quotations and original sources. The student keeps a notebook of names and ideas, and he is urged to read further in some appropriate source. In some instances biographies are suggested and students have reported to me that they were thoroughly inspired by these adventures. Physics takes on a more humanistic temper, and even the "technical" people enjoy it more. I show portraits and pictures of famous physicists and mathematicians and give brief biographical accounts. An elegant array of portraits is available from Scripta Mathematica with biographical accounts by Henry Crew. *Cenco News Chats* covers bear beautiful plates and brief biographical sketches. In fact, I have mounted a goodly number of these on masonite boards, shellacked them over with clear shellac, and ornamented my classroom and laboratory walls. The attention they received is surprising.

There is much indeed to be said for the historical and the humanistic in a first physics course, and we need to inject more of it.

JULIUS SUMNER MILLER

Dillard University
New Orleans, Louisiana

Impact Problems

OFTEN the author of an elementary physics textbook must introduce one or more simplifying assumptions in order that the student can solve a particular problem. Usually the simplifying assumptions have little effect on the physical reasonableness of the numerical result. Sometimes, however, this is not the case and the student gains a distorted impression of the real physical situation. He may retain this impression long after he has completed the course. Recently I came across what may be an example of such an occurrence and believe it worth calling to the attention of authors of physics textbooks.

A common belief is that a high velocity particle completely vaporizes on impact. For a number of years I have been studying phenomena associated with the impacts of high velocity missiles and find that what actually occurs is that most of the energy of the missile is transferred to the target rather than absorbed by the missile itself. Even for impacting velocities of 20,000 ft/sec the missile will often remain intact with little evidence of melting although, had an appreciable part of the kinetic energy of the missile been absorbed by it, the missile would certainly have melted and perhaps vaporized.

From whence, then, has the impression arisen that such a missile will vaporize? In attempting to track down the origin of this belief, I have examined a number of elementary physics textbooks and feel that it may be traceable to college physics courses. At any rate, the bullet-impact problems given in most texts are not good representations of real physical situations. For example, Duff[1] gives the following problem:

> A lead bullet of 50 g mass is fired into a target with a velocity of 50 meters/sec. If the heat produced goes into the bullet, how much is its temperature raised?

Sears[2] and Taylor[3] give similar problems. On the other hand Stewart[4] assumes that one-third of the energy and Foley[5] and Black[6] assume that one-half of the energy goes into heating the ball. Some authors, Eldridge[7] and Michener,[8] avoid the problem. Perkins[9] gives a very ambiguous problem:

> A bullet weighing 6 oz and moving with a speed of 1800 ft/sec strikes a target which stops it completely. How much heat is developed?

At least one author, Saunders,[10] substitutes an excellent problem:

> A "shooting star" is a fragment of rock or ore, traveling at a speed greater than that of any bullet, perhaps 40 km/sec. How does its kinetic energy compare in amount with the heat energy that can be obtained by burning an equal weight of coal?

Each of the above problems is a good one from the standpoint of fixing the concept of mechanical equivalence of heat. Should we do this, however, at the expense of giving the student false notions regarding the physical behavior of materials under impact?

JOHN S. RINEHART

Naval Ordnance Test Station
China Lake, California

[1] Duff, *Physics for students of science and engineering* (Blakiston, ed. 8, 1937), problem 24, p. 690.
[2] Sears, *Principles of physics I* (Addison-Wesley Press, 1947), problem 7, p. 374.
[3] Taylor, *Physics, the pioneer science* (Houghton Mifflin, 1941), problem 7, p. 303.
[4] Stewart, *Physics* (Ginn and Co., 1931), problem 12, p. 263.
[5] Foley, *College physics* (Blakiston, ed. 2, 1937), problems 1–6, p. 323.
[6] Black, *An introductory course in college physics* (Macmillan, rev. ed., 1941), problems 9 and 10, p. 312.
[7] Eldridge, *College physics* (Wiley, 1940).
[8] Michener, *Physics for students of science and engineering* (Wiley, 1947).
[9] Perkins, *College physics* (Prentice-Hall, ed. 3, 1948), problem 2, p. 221.
[10] Saunders, *A survey of physics* (Henry Holt, rev. ed., 1936), problem 1, p. 242.

No amount of experimentation can ever prove me right; a single experiment may at any time prove me wrong.—A. EINSTEIN (?)

ANNOUNCEMENTS AND NEWS

Book Reviews

Scientific Autobiography and Other Papers. MAX PLANCK. Pp. 187. Philosophical Library, New York, 1949. Price $3.75.

These are the last writings of Max Planck and as such would naturally command the interest of a wide circle of readers. Though the book is small (not much over thirty-five thousand words) it ranges over some of the most fundamental problems of science and philosophy to that of religion. In spite of the drab title it is the warm and human document of a man who has done and seen much, and who has had time in his last years (he reached nearly ninety) to look back and try to think things through. The professional philosopher may be disconcerted at times by contradictions in viewpoint, but Planck has labored to make sense out of a world in which contradictions were already present and even in his last years he has made progress.

In the "life" (thirty-one pages) Planck tells of his decision to devote himself to theoretical physics when the latter was only beginning to be thought of as a separate field. He recalls his student days, his teachers, and his early struggles for recognition. Following in the footsteps of Clausius he clarified the Second Law of Thermodynamics and embodied it in his doctoral dissertation at Munich in 1879. But he never forgot his chagrin when Kirchhoff disagreed with him, when Helmholtz (as he believed) did not even bother to read his paper and when Clausius himself, for some odd reason, failed even to answer his letters. However, recognition could not always be withheld, and in 1889, after the death of Helmholtz, Planck left his associate professorship in Kiel to take the chair at Berlin occupied by his former teacher.

One is impressed that Planck was moving steadily forward in spite of difficulties and lack of recognition until at last he reached the quantum theory. Even then he was reluctant to give up classical physics. Only after years of futile attempts at reconciliation did he admit the need for "totally new methods of analysis," and this same reluctance colors his philosophy.

In the second chapter Planck discusses phantom problems with which we are more familiar under the name "meaningless problems." As Planck points out, such problems are more common than is usually suspected, but in many cases the meaninglessness is not absolute but dependent. Herein lies one of the limitations of a radical positivism or operationalism, but Planck has come gradually to see much of value in these more modern viewpoints without becoming a wholehearted disciple. To ask whether an electron is a wave or a particle without defining the conditions of reference is to state a phantom problem and Planck carries this general idea to the difference between body and mind, and to the problem of causality *versus* indeterminism, and even to ethics. He disposes of the body-mind problem by concluding that "They are the self-same processes, only viewed from two diametrically opposite directions." This disposal of mind as a separate entity is not out of harmony with the recent views of certain radical positivists, though arrived at by a different path.

In the next two chapters Planck discusses the meaning and limits of exact science and the meaning of reality, and the dilemma of causality *versus* indeterminism. Like the earlier positivists Mach and Pearson, he recognizes the basis of science in sense experience, but where they would stop he goes on to assert the existence of an absolute reality.

Of the horns of the dilemma he takes the view that not only is causality not ruled out, but it is "as strictly valid" in quantum mechanics as in classical physics. However, the question still remains—how true is it? Returning now to a more positivist viewpoint he concludes that "causality is neither true nor false" but that it is a guidepost to help us in blazing new trails.

In his final chapter Planck finds no conflict between science and religion. Even positivism, he asserts, must have faith in something if only in the repeatability of physical measurements. But here he rejects positivism as barren and asserts his faith in an ultimate reality, in a rational order which we can more and more nearly come to know, and in God. Perhaps some of this reflects those bitter sufferings of his last years, in which he expressed the faith that permanent happiness comes more than anything else from "integrity of soul, which manifests itself in a conscientious performance of one's duty." The volume is prefaced by the memorial address delivered by Planck's former pupil, Max von Laue, in Göttingen, October, 1947.

ROGERS D. RUSK
Mount Holyoke College

Electrical Resistance Strain Gauges. W. B. DOBIE AND P. C. G. ISAAC. Pp. 114+xiv. English Universities Press, Ltd., London (Macmillan, New York), 1948. Price $3.50.

This little book is an excellent survey of the theory and applications of the electric resistance strain gauge. Oddly enough it is an English book on a recent American "gadget," based on a fundamental English discovery, and as stated in the preface, the book does have a distinctly "American flavour in places."

The book introduces the subject historically with the work of Lord Kelvin on the change in electrical resistance of metal wires with strain. The later quantitative work on strain-resistance relationships of various wires is introduced and summarized. This is a convenient reference of the properties of the various types of strain gauges not found elsewhere in handbooks, etc. The book continues with descriptions of the fabrication of the grid-type gauges, and a detailed treatment of the methods and materials used in bonding the grids to the subject under test. There

is a chapter on the determination of static strains and one on dynamic strains with rather detailed descriptions of available commercial amplifiers, indicators, and recorders.

Nearly half of the book is given over to a rather elementary treatment of stress analysis and electronics as a necessary adjunct to the successful use of the strain gauges. Twenty-three pages are devoted to a review of strength of materials, determination of principal stresses, Mohr circles, etc., while thirty-four pages are devoted largely to cathode-ray oscillographs, electronics, fundamental bridge circuits, etc. It is obvious that the success of a gauge depends upon the accuracy with which the change in resistance can be measured, so that some discussion of special bridges is desirable. However, this reviewer was disappointed to find so much of the book devoted to such well-known phenomena as the rectifier and oscillating circuits and a deficiency in actual working circuits (not schematics) for experimenters and users of strain gauges.

The chapter on applications is interesting and could well have been extended. The authors point out that the last chapter on brittle lacquers has nothing to do with strain gauges, but was closely related to the general problem of stress analysis. Except for the fact that photoelastic methods are omitted, the authors could have more aptly entitled the book "Elementary Stress Analysis" and included some of the more recent work on stress concentrations, residual stresses, etc. Howeve^r, the book is a contribution to the whole subject and is most of all a handy source of references and information about resistance strain gauges not found together elsewhere.

GRANT O. GALE
Grinnell College

New Members of the Association

The following persons have been made members or junior members (*J*) of the American Association of Physics Teachers since the publication of the preceding list [*Am. J. Physics* **17**, 558 (1949)].

Abramson, Adolf Avraham (*J*), 630 W. 135th St., New York 31, N. Y.
Ananikian, Vahan, 231 Sigourney St., Hartford, Conn.
Barricklow, Charles H. Jr. (*J*), 523 E. Highland Ave., Ravenna, Ohio.
Beckman, Norman John (*J*), 20 Southworth St., Williamstown, Mass.
Benedict, Robert Neil (*J*), 29 Darlington Rd., Delaware, Ohio.
Biersdorf, William Richard (*J*), 806 Colorado St., Pullman, Wash.
Birkhoff, Robert D., University of Tennessee, Knoxville, Tenn.
Bolgiano, Louis Paul Jr., Apt. A, 5600 Abner Ave., Baltimore 12, Md.
Brown, Edmond Joseph, 2710 Kittrell Dr., Raleigh, N. C.
Burden, Henry S. Jr. (*J*), P.O. Box 616, Summerville, S. C.
Burke, Edward Walter Jr., King College, Bristol, Tenn.
Campos, Alberto Manuel Jr., 31 Twelfth St., La Loma, Quezon City, Philippine Islands.
Davis, Charlton Edward (*J*), Williams College, Williamstown, Mass.
Denny, Paul, East Central State College, Ada, Okla.
Dobriansky, Bohdan John (*J*), 325 East 16th St., New York 3, N. Y.
Doran, Ray Lewis, University of Utah, Salt Lake City, Utah.
Estes, Omar C., Henderson State Teachers College, Arkadelphia, Ark.
Gaeddert, Willard, 3926 Baldwin Ave., Lincoln, Neb.
Gallup, Kenneth W. (*J*), 2659-C Trinity Dr., Los Alamos, N. M.
Godwin, John Cochran, 56 College Ave., Buckhannon, W. Va.
Greene, Lawrence C., 116 E. Hammond St., Otsego, Mich.
Gregory, Richard Leland (*J*), Box 301, Attica, Kan.
Guilford, Edward C., 321 W. First St., Chico, Calif.
Hamilton, Robert J., 559 Tenth Ave., New Brighton, Pa.
Hotopp, Alfred H., 45 Hillside Ave., Caldwell, N. J.
Hudson, Alvin M. (*J*), Stanford University, Stanford University, Calif.
Hultquist, Paul Frederick, 2403 Arapahoe St., Boulder, Colo.
Kennedy, Margaret M., 2280 Andrews Ave., New York 53, N. Y.
Lillestrand, Robert L. (*J*), Pioneer Hall, University of Minnesota, Minneapolis, Minn.
McLeod, Norman J. (*J*), Route 1, Walla Walla, Wash.
Martin, Gruver Howard (*J*), University of North Carolina, Chapel Hill, N. C.
Mohr, Eugene Irving, Box 34, Collegedale, Tenn.
Mouw, Ralph J., 206 S. Williams St., Orange City, Iowa.
Niedner, Jack Bernard (*J*), 7464 York Dr., Clayton 5, Mo.
Numbers, Richard Scott (*J*), 775 B. Quartz St., Los Alamos, N. M.
Olson, Keith W. (*J*), 3104 Lincoln Way, Ames, Iowa.
Page, Allen Donald, 146 N. Sleight St., Naperville, Ill.
Riggins, Ira Dale, Washington State College, Pullman, Wash.
Schweighardt, Robert John (*J*), 8425 Belford Ave., Los Angeles 45, Calif.
Squires, Burton Elliott, Jr., 114 South Atherton St., State College, Pa.
Vedder, James F. (*J*), Room 468, International House, Berkeley 4, Calif.
Wajda, Edward Stanley, 633 Lansing St., Schenectady 3, N. Y.
Walker, John Baldwin, Seminary St., New Canaan, Conn.
Webster, Williams Kendrick (*J*), Annex Hall, University of Illinois, Urbana, Ill.
Weeks, Walter L., Michigan State College, East Lansing, Mich.
Woessner, Russell Howard (*J*), Box 56, Fitch Rd., Clinton, Mass.
Wolf, Bryant Edward (*J*), Washington University, St. Louis 5, Mo.